Molecular Pathology

for

Toxicologic Pathologists

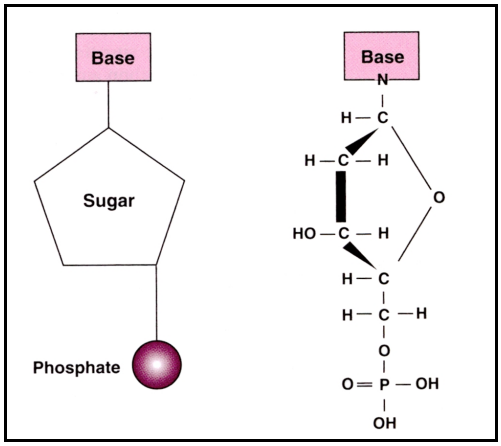
A Continuing Education Course
1999

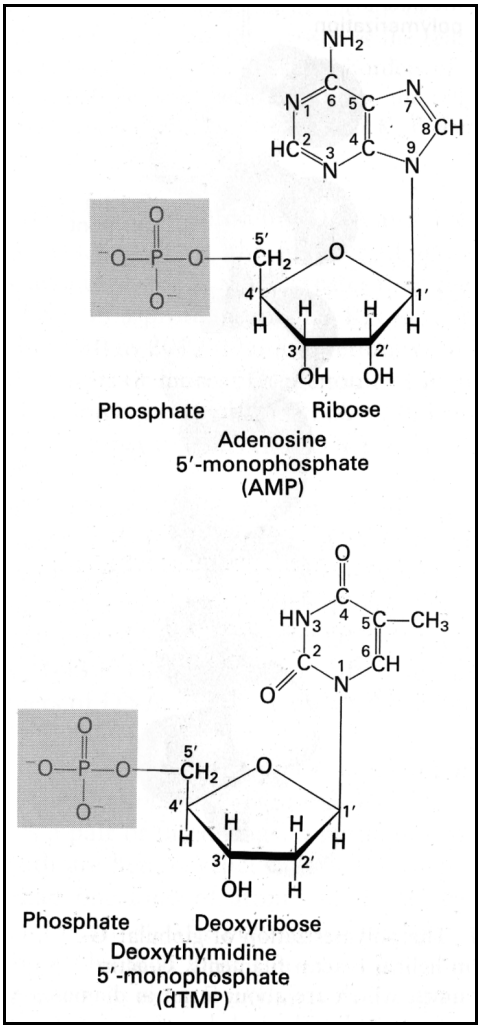
CONTENTS

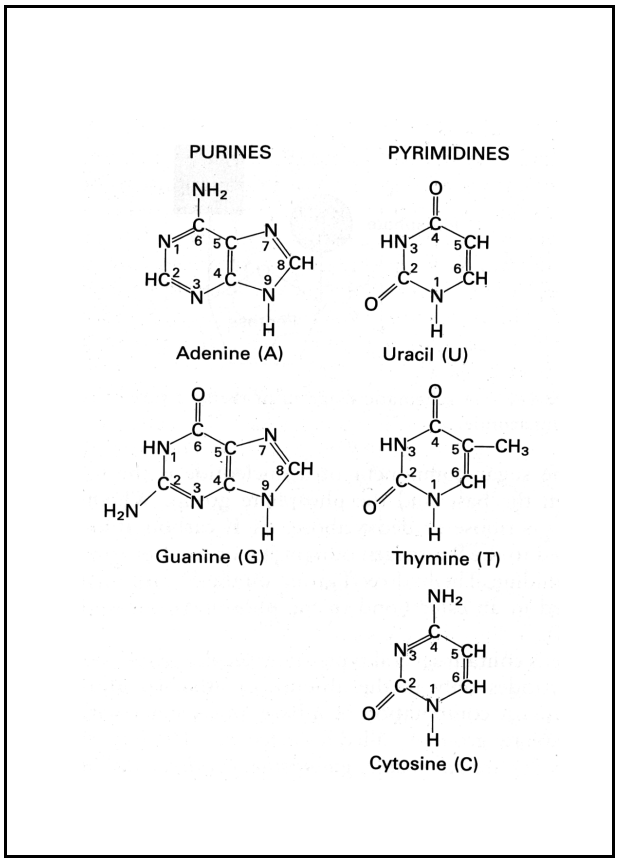
NUCLEIC ACIDS

[Figures from: D.P. Clark and L.D. Russell (1997) *Molecular Biology Made Fun and Simple* with permission from Cache River Press, Vienna, IIllinois, USA.]

Nucleic acids are comprised of chemically linked sequences of **nucleotides**. Each nucleotide is comprised of a cyclical nitrogen-containing **base**, a 5-carbon **sugar**, and a **phosphate** group. It is the sequences of the bases that constitutes the genetic information encoded in the genome.



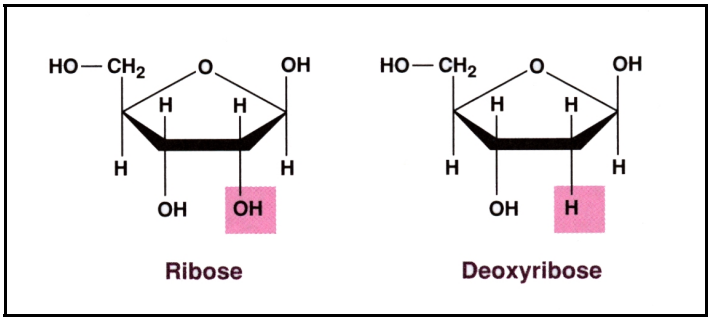




There are two types of bases - **purines** (adenine and guanine) and **pyrimidines** (cytosine and thymine in DNA and cytosine and uracil in RNA). Purines are fused 5- and 6-member rings and

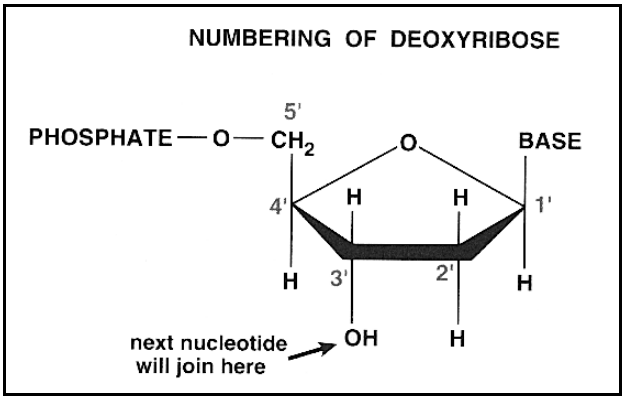
pyrimidines are 6-member rings. These bases constitute the genetic alphabet and consist of the letter abbreviations A, T, C, and G in DNA and A, U, C, and G in RNA.

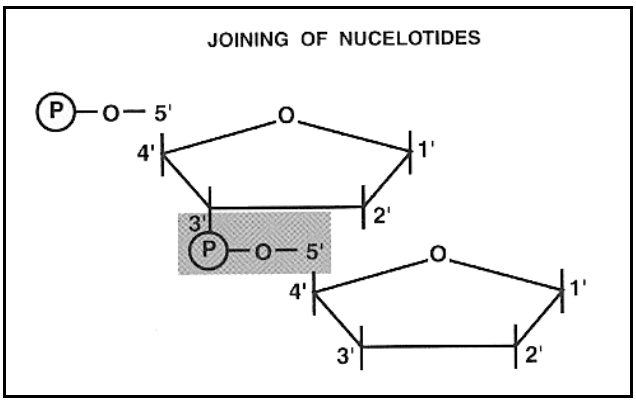
When a 5-carbon sugar (pentose) is linked to a purine or a pyrimidine, it is called a **nucleoside**. The linkage to the sugar is from the N_1 position of the pyrimidine or the N_9 position of the purine to the C_1 position of the sugar. Since both the nucleotides and the sugars are numbered, the positions of the carbons of the sugars are conventionally designated as prime (’). The sugar (pentose) that is present in DNA is deoxyribose and the sugar in RNA is ribose. The only difference between ribose and deoxyribose is the presence of a hydroxyl group on the 2’ position of the sugar.

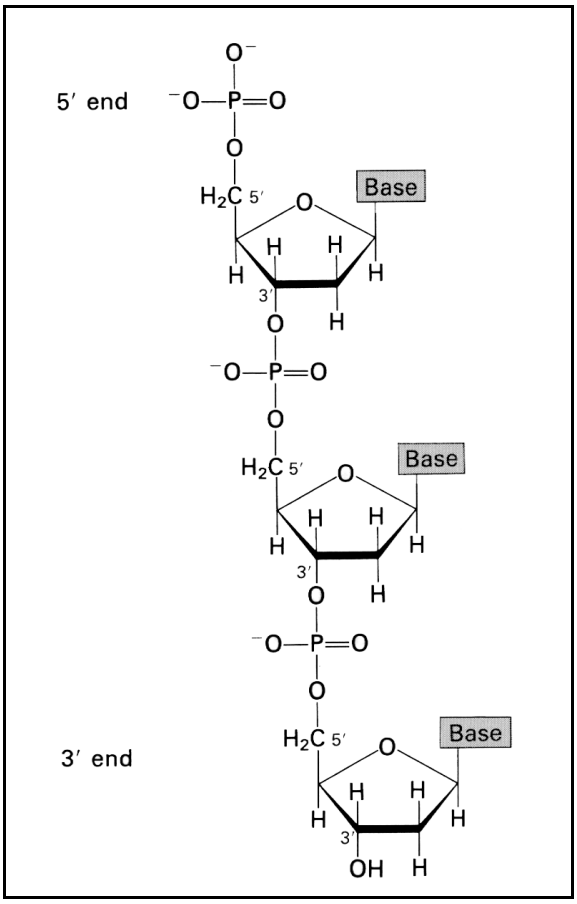


Because of the OH group of ribose, RNA is less chemically stable than DNA. RNA is thus considered base-labile.

The nucleotides (a combination of a nitrogenous base, a sugar, and a phosphate group) connect together to form a nucleic acid chain following a specific pattern of connection. The 5’ position of one sugar ring is connected to the 3’ position of the next sugar ring by a phosphate group. Thus, each nucleic acid chain starts with a free 5’ phosphate on the sugar and ends with a 3’ hydroxyl on the sugar. This 5’ and 3’ arrangement is reflective of the basic molecular concept of polarity (the 5’ to 3’ concept).





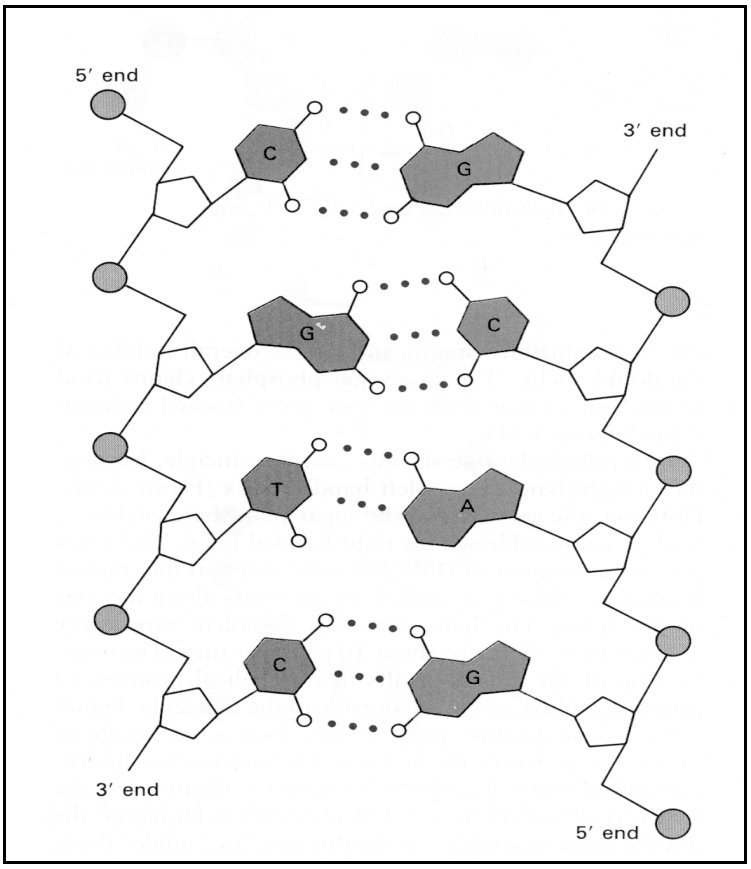


The following table sorts out the relevant nomenclature:

Base	Nucleoside	Nucleotide	Abbreviations	
			RNA	DNA
Purine				
Adenine	Adenosine	Adenylic acid	AMP	dAMP
Guanine	Guanosine	Guanlyic acid	GMP	dGMP
Pyrimidine				
Cytosine	Cytidine	Cytidylic acid	CMP	dCMP
Thymidine	Thymidine	Thymidylic acid		dTMP
Uracil	Uridine	Uridylic acid	UMP	

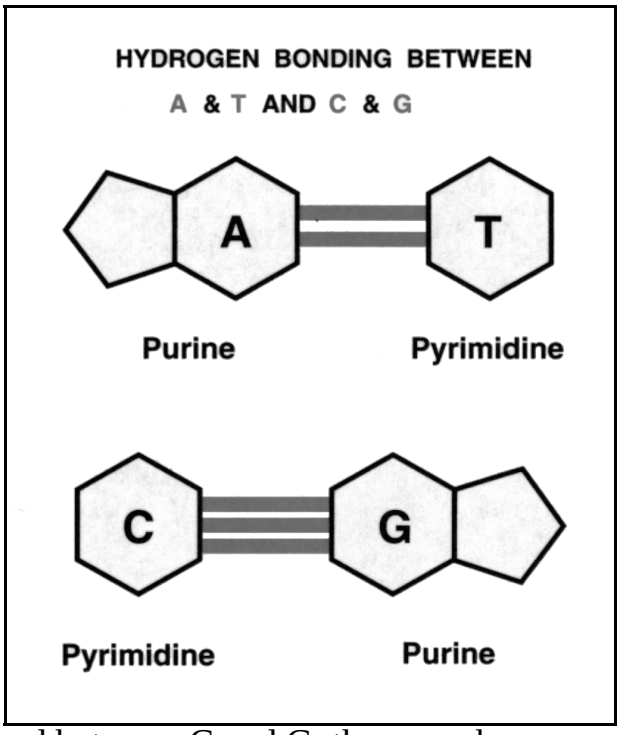
While the above nomenclature may not be of particular concern to pathologists, it is of some interest when considering two common agents that are used to label proliferating cells - **tritiated thymidine and bromodeoxyuridine**. It can be seen that these are named after the related nucleoside and, in fact, these agents are

referred to as thymidine analogs and are incorporated into newly synthesized DNA in place of thymidine.

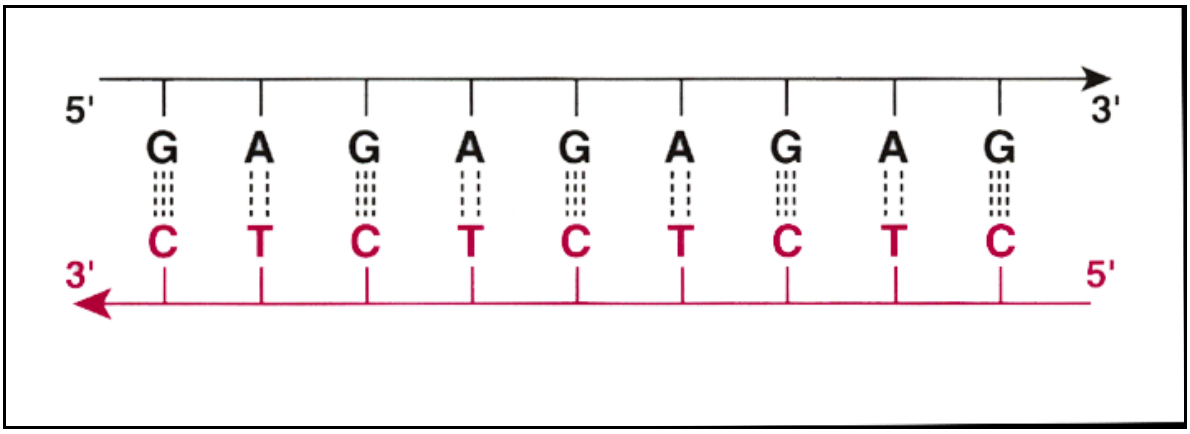


Another basic concept of molecular biology is **complementarity**. Within the double-stranded DNA helix, a purine from one strand of DNA is always paired with a pyrimidine from the other strand. The two DNA strands are connected to one another via hydrogen bonds which form between the bases.

A & G are purines while C & T are pyrimidines. In a point mutation called a **transversion** a purine is replaced by a pyrimidine or vice versa. In a **transition**, a purine is replaced by a purine or a pyrimidine is replaced by a pyrimidine.



Adenine (A) always pairs with thymidine (T) via two hydrogen bonds while guanine (G) always pairs with cytosine (C) via three hydrogen bonds. The A-T and G-C base pairs are spoken of as being complementary. As a consequence of this specific pairing in DNA, the two strands are always equidistant between the larger purine molecule and the smaller pyrimidine molecule. The fact that there are 3 hydrogen bonds between G and C and only 2 between A and T explains why it takes more energy to separate a GC pair than to separate an AT pair. Thus, GC-rich sequences in the DNA are relatively more stable than AT-rich sequences. Because of the complementarity between A and T and between C and G, there can be accurate reproduction of genetic material as each stand can serve as a template for the synthesis of the opposite strand. Because of the complementarity between the two strands of DNA, they run in opposite directions from one another and are referred to as being **antiparallel** to each other.



DNA Methylation

The cytosine of DNA is sometimes methylated at the C5 position to form methylcytosine. In mammals, the sites of methylation are typically cytosine bases that are followed by a guanine on the same strand. This is sometimes written as

CpG. There are two noteworthy points about DNA methylation. First, methylation on a gene, especially heavy methylation, results in reduced expression of that gene. Distinct patterns of DNA methylation are found in different tissues. Second, the pattern of methylation is passed on when DNA replicates by the action of DNA methylase. This represents a heritable epigenetic change.

Genes that are constitutively expressed in all cells (housekeeping genes) typically exist in a permanently unmethylated state. The absence of the methyl groups probably provides a signal for their continued expression. Genes that show tissue-specific control have variations in DNA methylation so that only the genes appropriate for the function of that cell are unmethylated and, therefore, will be transcribed. The absence of methyl groups on DNA is believed to be associated with the ability to be transcribed but not with the act of transcription itself.

An important regulatory role of DNA methylation occurs in the phenomenon know as **genomic imprinting** in which either a maternal or paternal allele is methylated and, thus, silenced.

DNA versus RNA

In contrast to DNA, RNA is single-stranded and the pyrimidine uracil replaces the DNA base thymidine. The chemical difference between uracil and thymidine is the absence of a methyl group at position 5 on the ring in uracil. Another difference is that the pentose ribose is present in RNA as opposed to deoxyribose in DNA (see

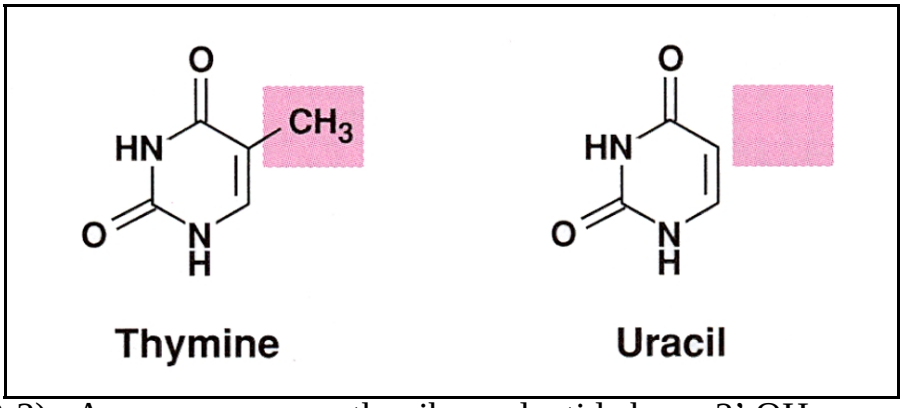


figure on 2-2). As a consequence the ribonucleotide has a 2' OH group on the sugar while this OH group is not present in DNA. This accounts for RNA being less chemically stable than DNA. Thus, RNA is spoken of as being base labile.

Types of RNA

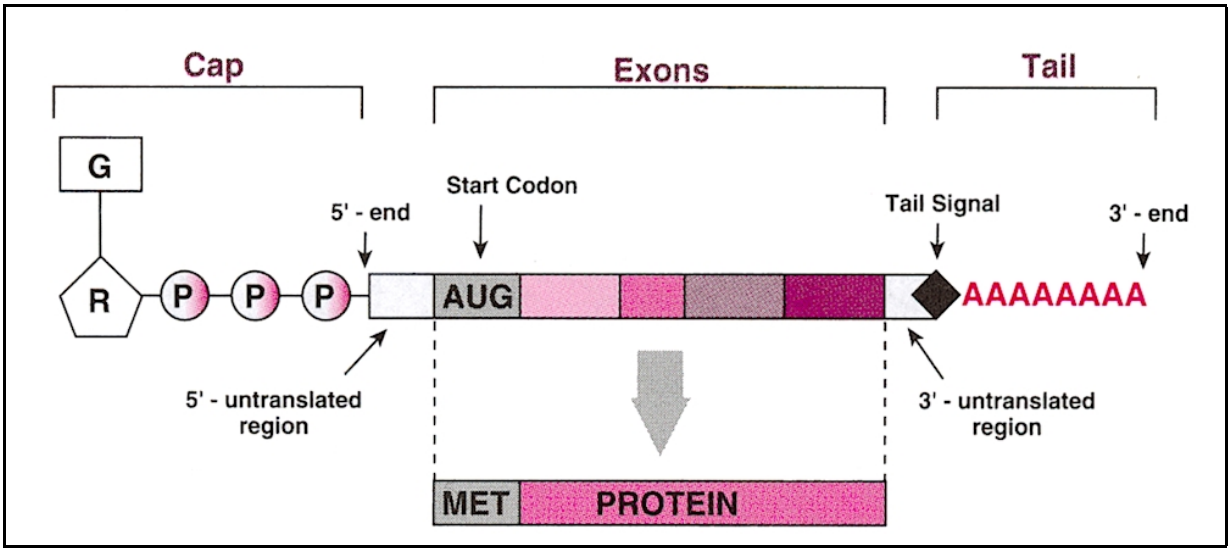
There are three types of RNA in the cell: messenger RNA (mRNA), ribosomal RNA (rRNA), and transfer RNA (tRNA). All three types are transcribed from DNA by a similar mechanism but a different enzyme (RNA polymerase) is involved with the synthesis of each type of RNA in eukaryotic cells.

Each run of three mRNA nucleotides is called a codon. Each codon encodes one amino acid. This can be thought of as letters of a 4 character alphabet spelling words that are each three letters in length. Stringing several words together makes a sentence (a string of adjacent amino acids whose sequence defines a specific protein). Thus, the genetic code is embodied in the specific codons. While some codons specify only one particular amino acid, other specific amino acids are encoded by more than one different codons. Because of this phenomenon, the genetic code is sometime referred to as degenerate (a better term might have been redundant). In addition there are three codons which are nonsense or termination codons and these function to terminate translation.

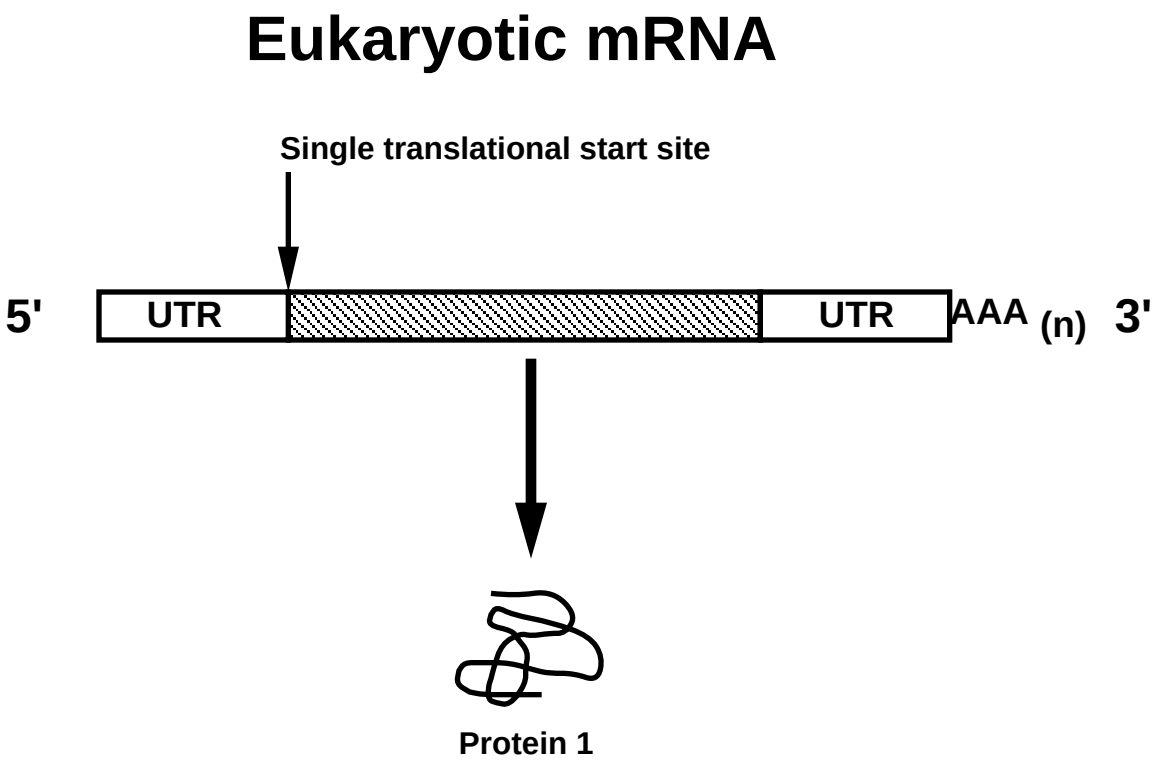
The genetic code is specified in the table below.

1st Base	2nd (middle) Base: U	C	A	G	3rd Base
U	UUU Phe	UCU Phe	UAU Tyr	UGU Cys	U
	UUC Phe	UCC Phe	UAC Tyr	UGC Cys	C
	UUA Leu	UCA Leu	UAA STOP	UGA STOP	A
	UUG Leu	UCG Leu	UAG STOP	UGG Trp	G
C	CUU Leu	CCU Pro	CAU His	CGU Arg	U
	CUC Leu	CCC Pro	CAC His	CGC Arg	C
	CUA Leu	CCA Pro	CAA Gln	CGA Arg	A
	CUG Leu	CCG Pro	CAG Gln	CGG Arg	G
A	AUU Ile	ACU Thr	AAU Asn	AGU Ser	U
	AUC Ile	ACC Thr	AAC Asn	AGC Ser	C
	AUA Ile	ACA Thr	AAA Lys	AGA Arg	A
	AUG Mat	ACG Thr	AAG Lys	AGG Arg	G
G	GUU Val	GCU Ala	GAU Asp	GGU Gly	U
	GUC Val	GCC Ala	GAC Asp	GGC Gly	C
	GUA Val	GCA Ala	GAA Glu	GGA Gly	A
	GUG Val	GCG Ala	GAG Glu	GGG Gly	G

mRNA molecules typically have a cap at the 5’ end. This **cap** involves a methylated guanosine which is attached to the first nucleotide of the mRNA via a triphosphate group. In addition, there is a **poly(A) tail** consisting of 150 to 250 adenine nucleotides at the 3’ end of the mRNA. Neither the cap nor the poly (A) tail is encoded by the DNA, but rather these are added after the start of transcription. The presence of the poly(A) tail allows for separation of mRNA from tRNA and rRNA, neither of which has a poly(A) tail.



Many mRNA molecules also have sequences at the 5' (leader) and 3' (trailer) ends which are not translated into protein and are termed **UTR**'s for untranslated regions.



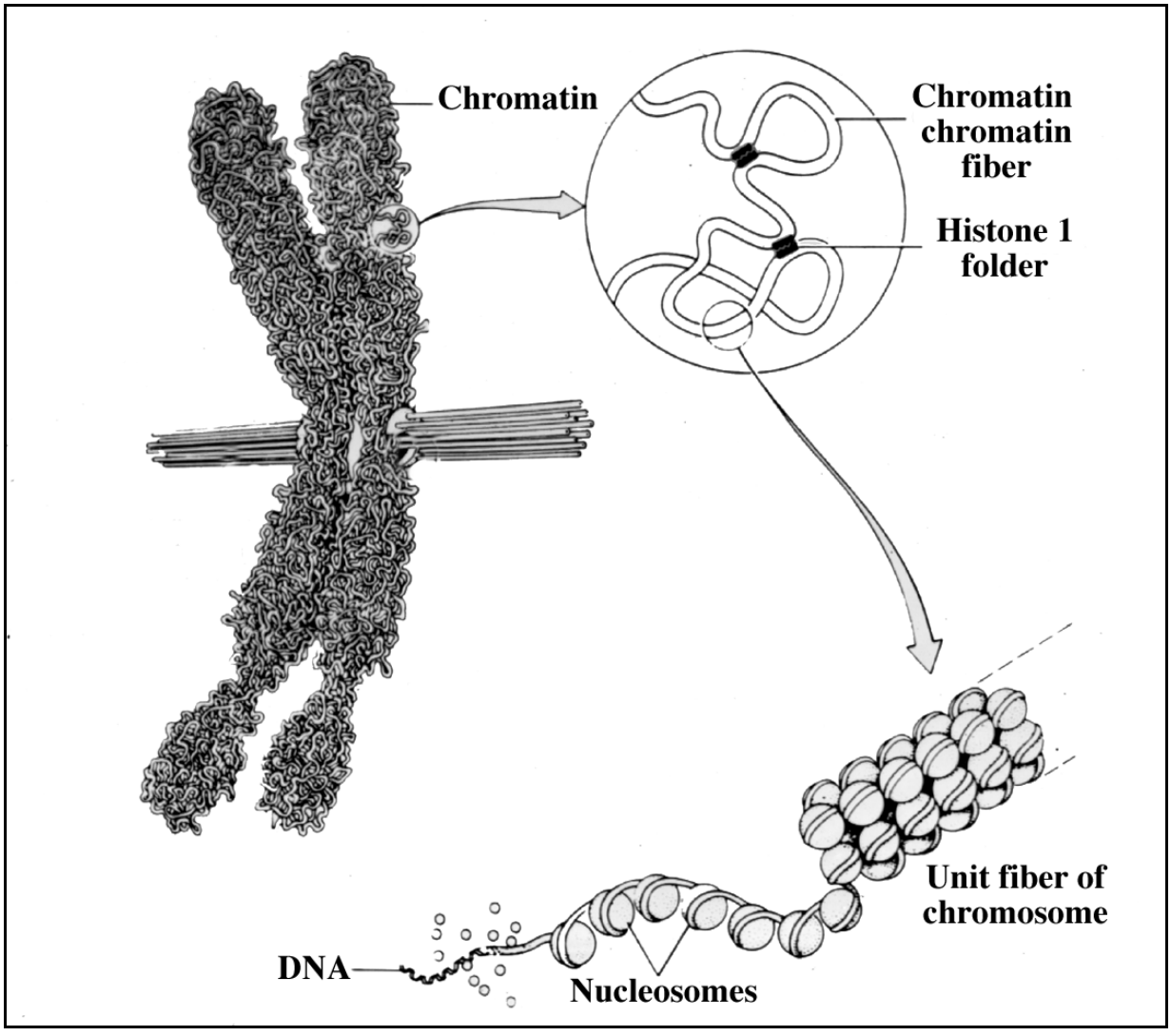
Most of the RNA in a cell is **rRNA** and is located in the ribosomes. rRNA does not code for proteins but helps decode the information contained in the mRNA.

There are many types of **tRNA** in the cell. Like the rRNA, tRNA also helps to decode mRNA but is not itself translated into protein. Each of the different tRNAs recognizes and binds to one of the 20 amino acids and interacts with mRNA by virtue of its unique three dimensional structure.

GENE STRUCTURE

Eukaryotic genes consist of sequences of DNA which are genuine coding sequences (exons) and intervening non-coding sequences (introns). In addition there are additional non-coding sequences of DNA between genes. It is estimated that only about 5% of the human genome contains actual coding sequences. Hundreds to thousands of copies of repetitive sequences occur both within genes, possibly representing multiple copies of highly used genes, and between genes, possibly serving a spacers. Their exact function is unknown. One commonly occurring highly repetitive sequence is the Alu sequence (Alu element) consisting of 300,000 to 500,000 base pairs scattered throughout the human genome.

DNA is tightly coiled (supercoiled) and packed in chromosomes. Chromatin refers to DNA sequences coiled around histone proteins.



Double stranded DNA sequences approximately 200 base pairs in length are wrapped around 9 histones forming a nucleosome. Chromatin actually consists of multiple nucleosomes.

The small single histone is H1 while the 8 clustered histones represent two copies each of H2A, H2B, H3 and H4.

During DNA replication, histones become accessible for in situ histone staining & this feature may be used as an indication of active cell proliferation.

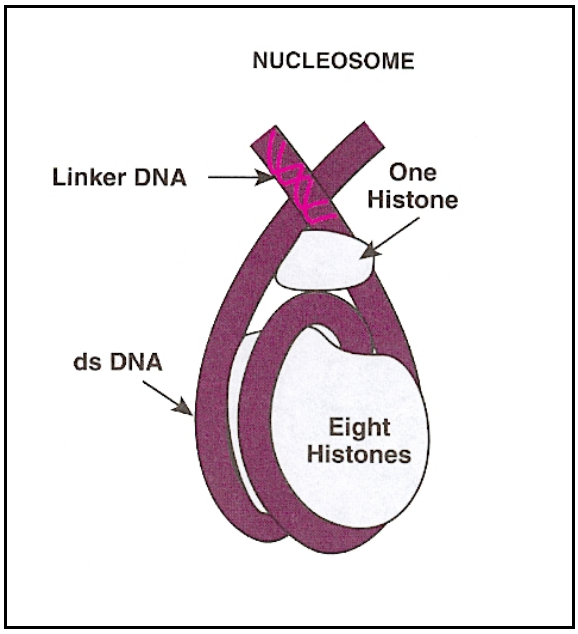
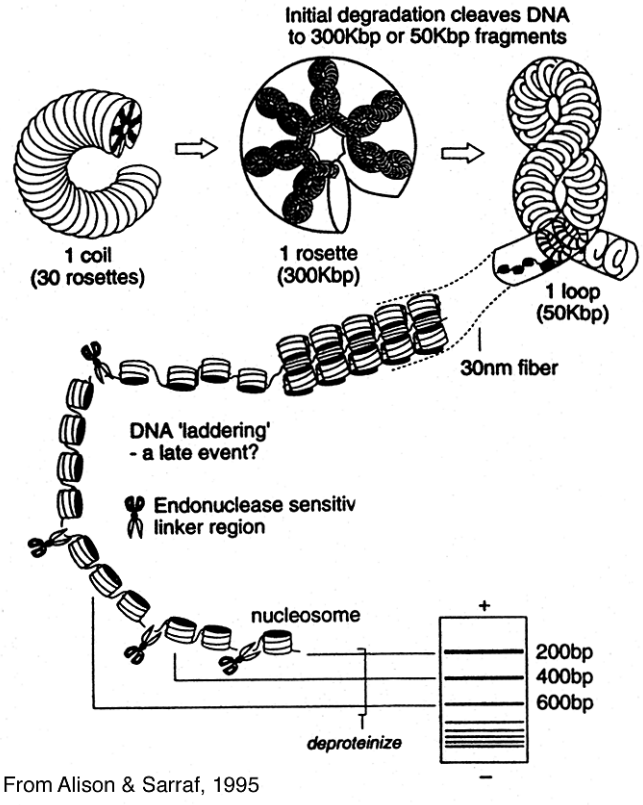


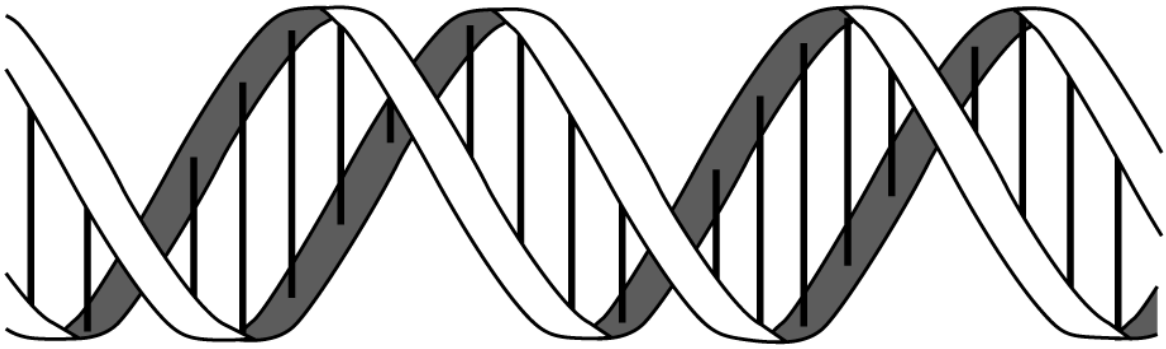
Figure from: D.P. Clark and L.D. Russell (1997) *Molecular Biology Made Fun and Simple* with permission from Cache River Press, Vienna, IIllinois, USA.

Internucleosomal cleavage of DNA occurs during the process of apoptosis and electrophoresis of the resulting DNA fragments produces the DNA ladder pattern that is a hallmark of apoptosis.

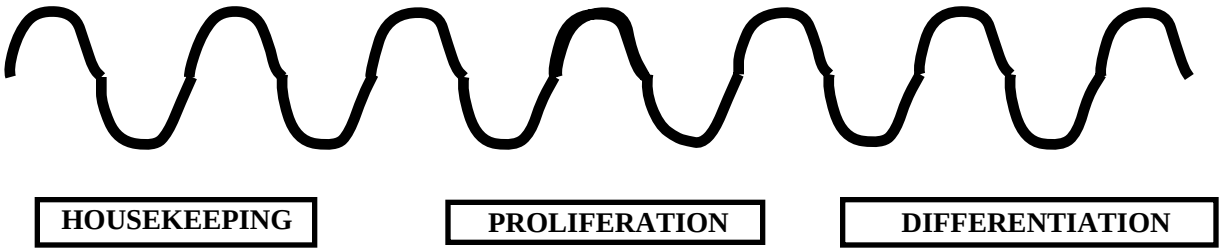


From Alison & Sarraf, 1995

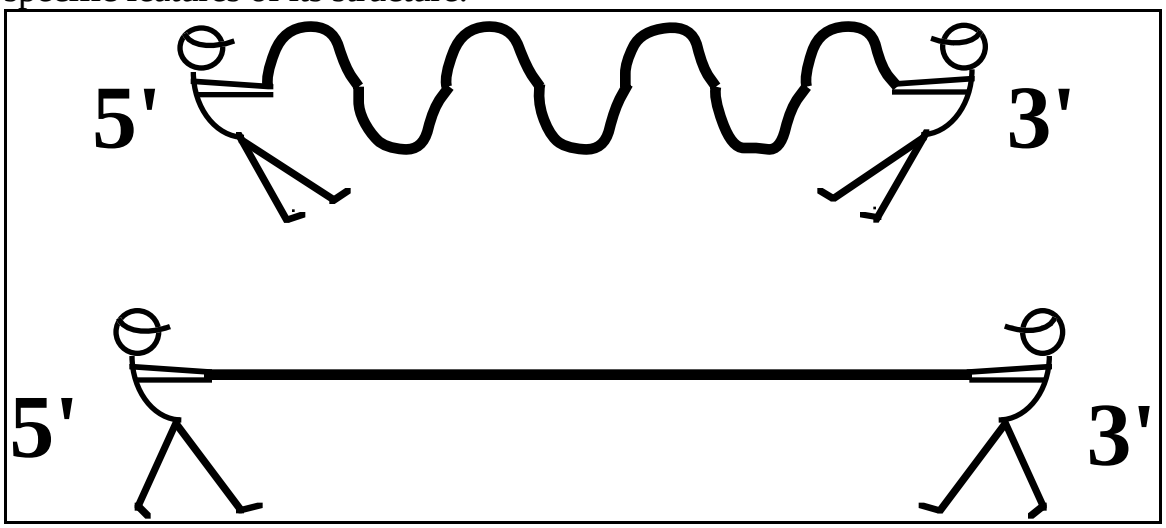
DNA exists as a double helix with the two parts joined by hydrogen bonds.

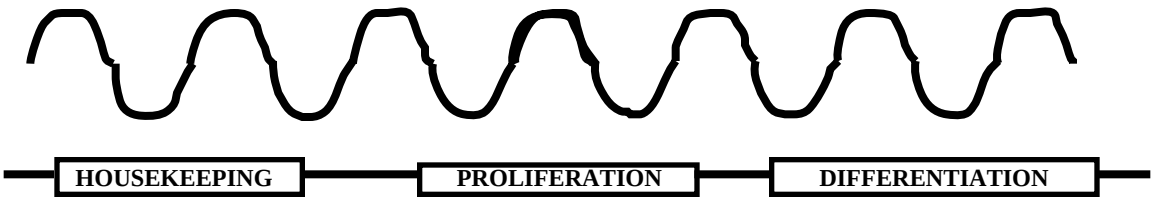


Different genes are arranged along the length of each strand of the double helix.



In conventional diagrams, DNA strands are straightened out to help illustrate specific features of its structure.





A gene includes regions that precede and follow a more centrally located **coding** region. Genes for polypeptides include a **leader** region followed by the coding region followed by the **trailer**.

Leader -----> coding region -----> trailer

Intervening sequences separate genes on a strand of DNA.

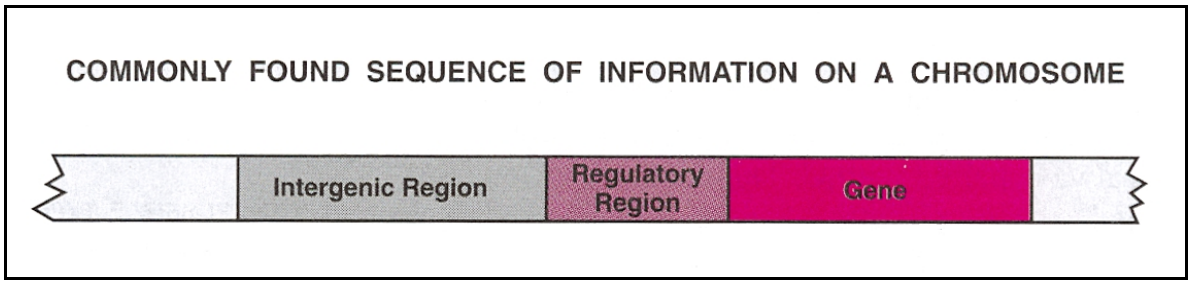


Figure from: D.P. Clark and L.D. Russell (1997) *Molecular Biology Made Fun and Simple* with permission from Cache River Press, Vienna, IIllinois, USA.

The leader and trailer are not translated into protein. The coding region is divided into exons and introns.

Leader -----> coding region -----> trailer

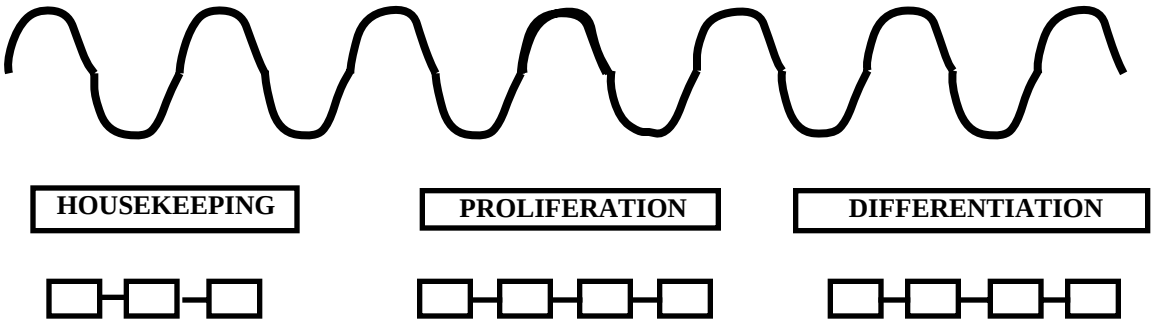
exons --- introns

Exons encode or **ex**press specific amino acids that ultimately make up the polypeptide gene product. **Introns int**errupt or **int**ervene between exons. While introns are transcribed into immature RNA, they are removed from

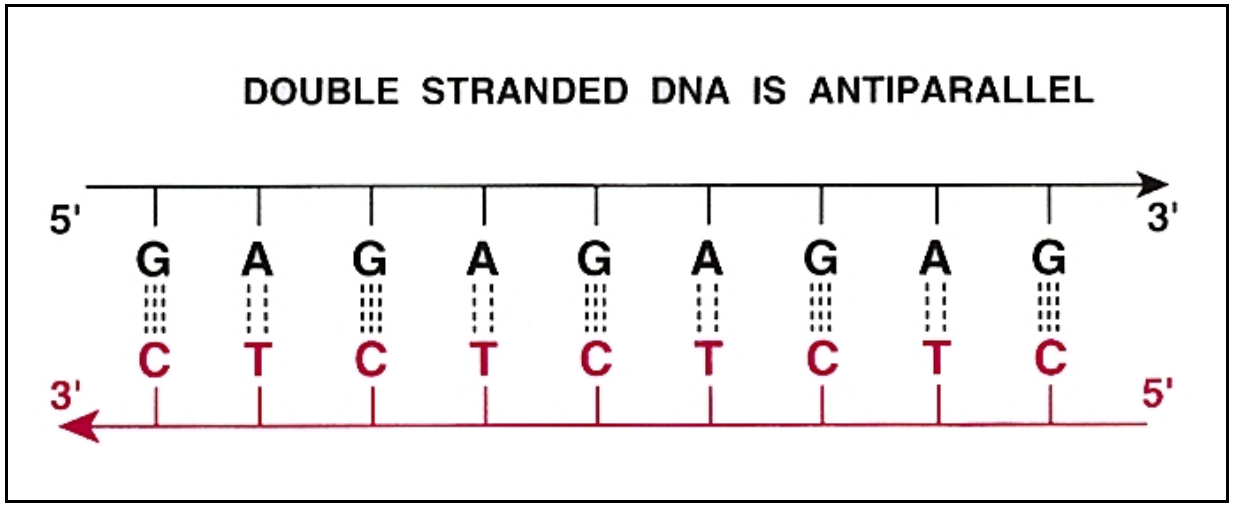
within the transcript by **splicing** together of exons on either side and, thus, do not encode amino acids.

Leader ----(intron---exon---intron---exon---intron)$_n$----trailer

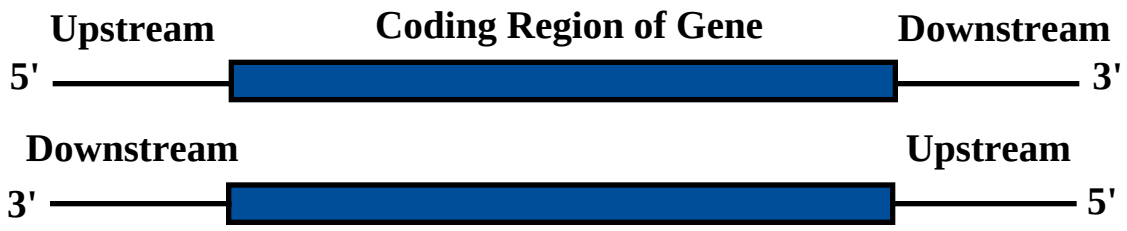
The separate strands of DNA are defined as a coding strand and a template strand.



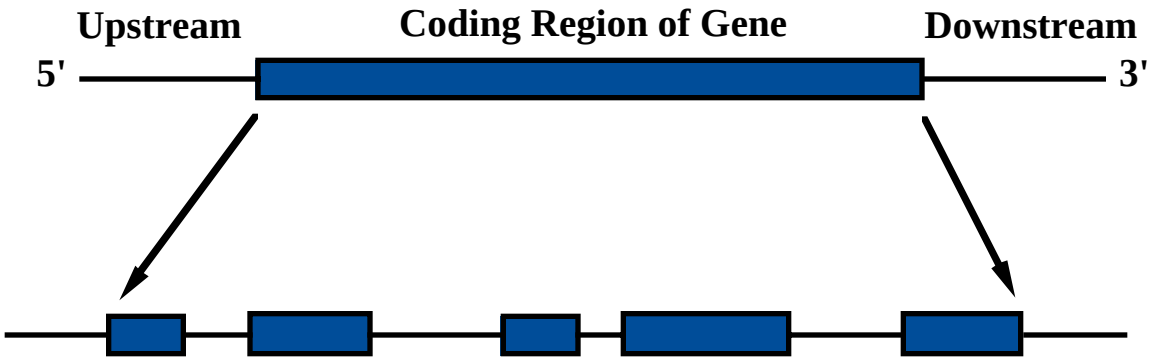
The genes have polarity characterized by a 5' upstream end and a 3' downstream end. In its double stranded form the complementary strands are antiparallel.



The following 3 illustrations are meant to summarize what has been presented relative to gene structure.



Template Region of Gene



= Exons **= Introns**

Exon = Expressed portion of the DNA
Intron = Intervening portion of the DNA

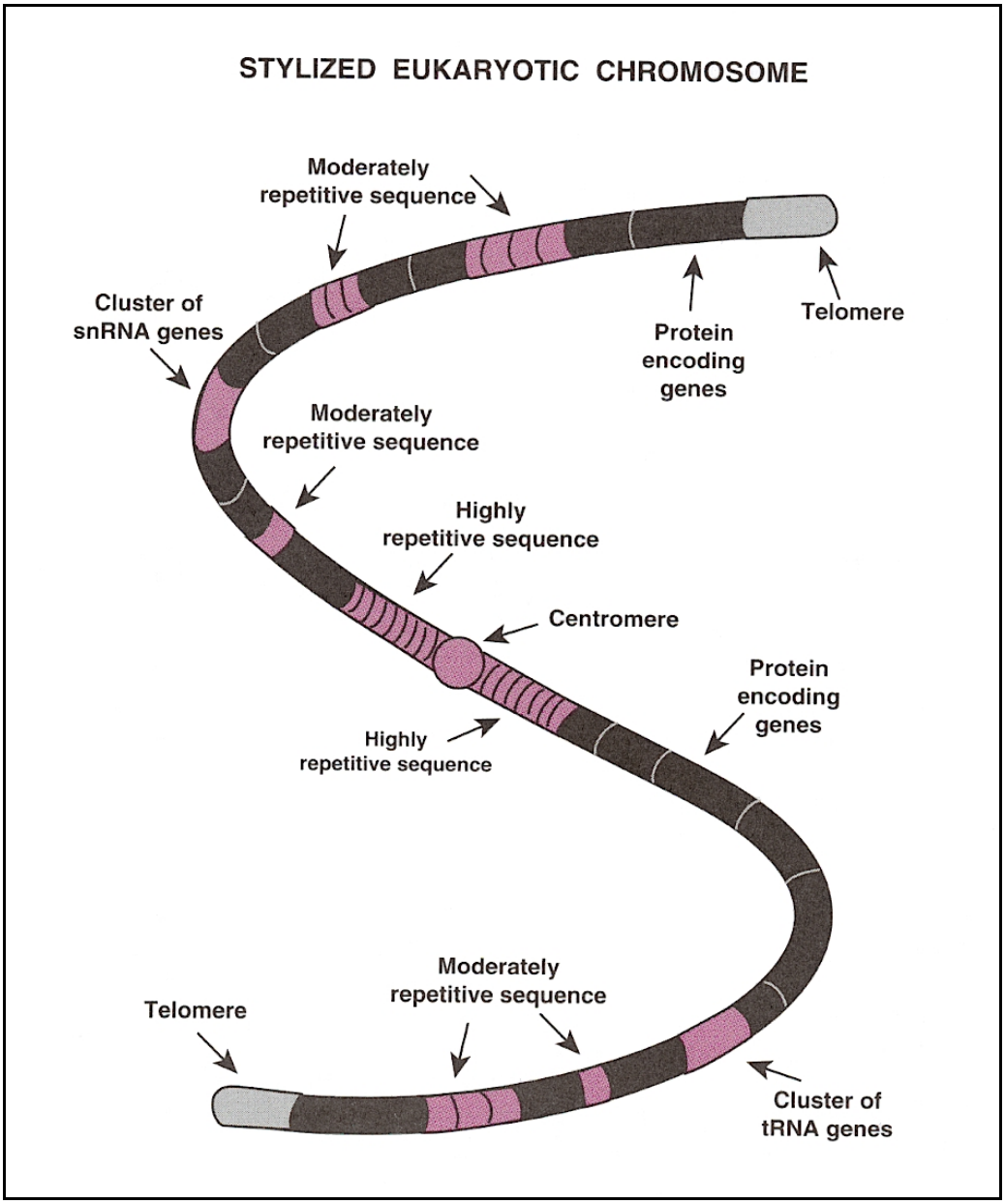


Figure from: D.P. Clark and L.D. Russell (1997) *Molecular Biology Made Fun and Simple* with permission from Cache River Press, Vienna, IIllinois, USA.

Repetitive sequences - Hundreds to thousands of copies of repetitive sequences occur both within genes, possibly representing multiple copies of highly used genes, and between genes, possibly serving a spacers. Their exact function is unknown. One commonly occurring highly repetitive sequence is the *Alu* sequence (*Alu* element) consisting of 300,000 to 500,000 base pairs scattered throughout the human genome. These contain a single site for the restriction endonuclease *Alu*I.

snRNA = small nuclear RNA; involved in splicing of mRNA.

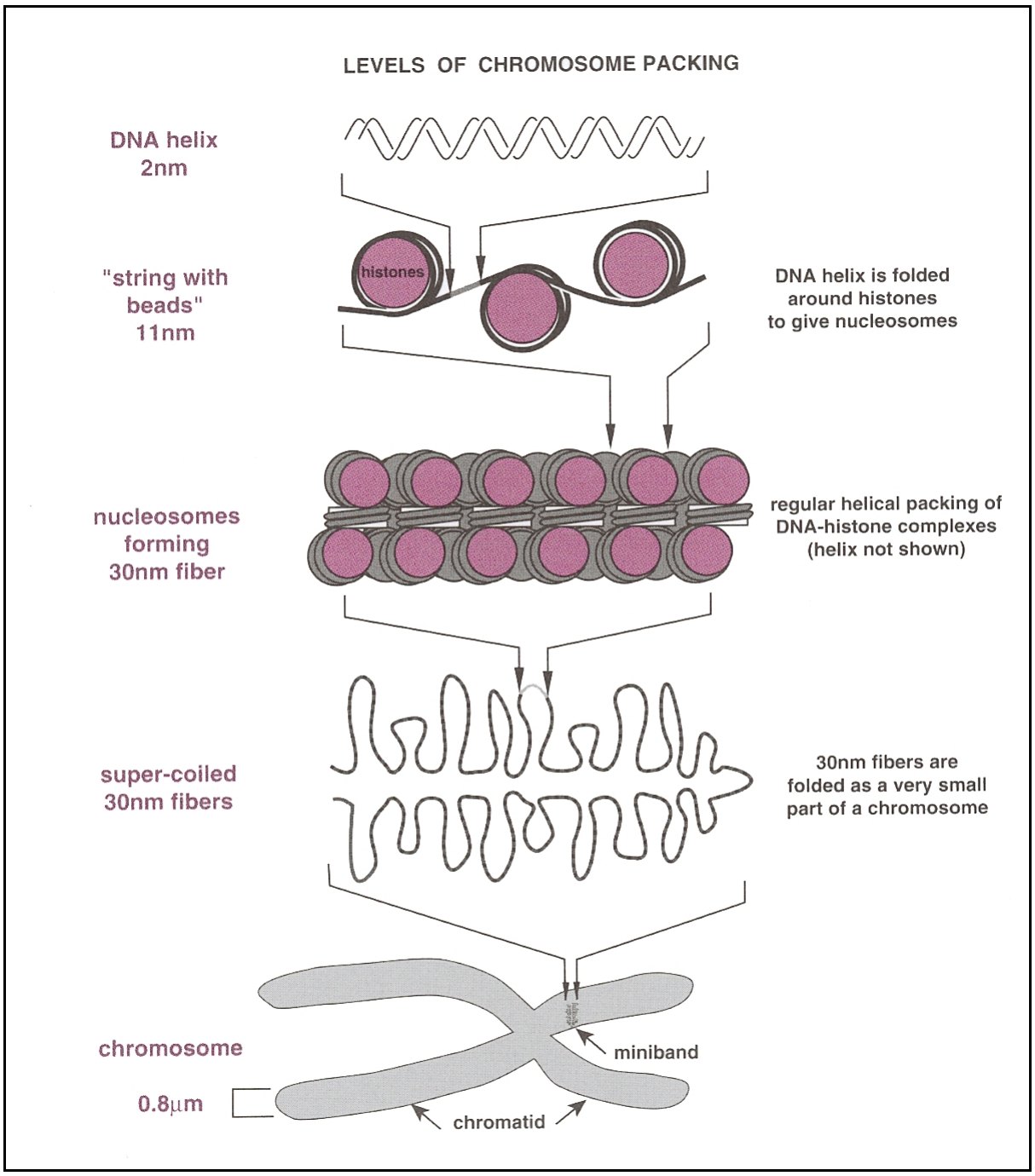


Figure from: D.P. Clark and L.D. Russell (1997) *Molecular Biology Made Fun and Simple* with permission from Cache River Press, Vienna, Illlinois, USA.

DNA REPLICATION

Replication is the process of making duplicate copies of DNA and it is based upon the principles of complementarity and base pairing. The term semi-conservative replication is used to indicate replication in which each daughter molecule of DNA gets one of the original DNA strands along with a newly synthesized complementary strand.

The supercoiled DNA must first be unwound and this is done by an enzyme called DNA gyrase. The DNA is still in a double helix and the double helix is unwound by the enzyme DNA helicase. The DNA helicase disrupts the hydrogen bonds holding the complementary strands together. It's action is localized and it does not result in breaking the DNA chain. Single strand binding protein (SSB) keeps the separated strands apart.

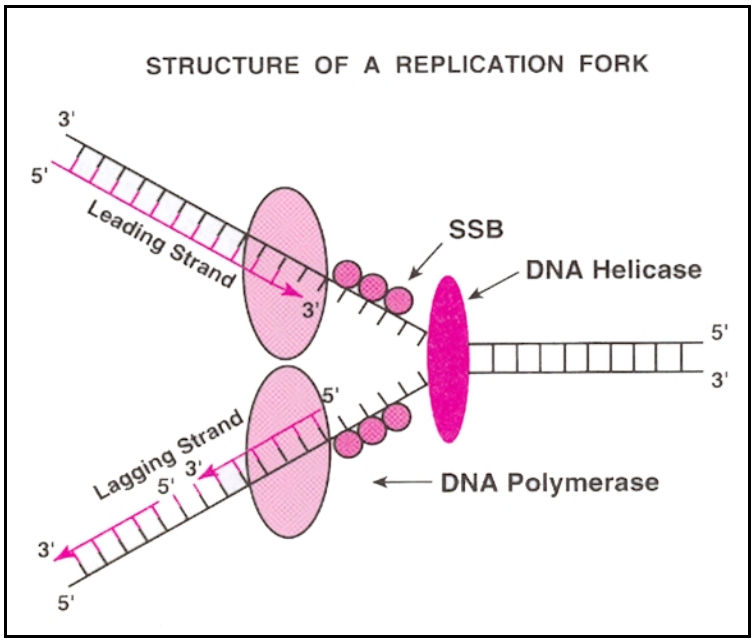
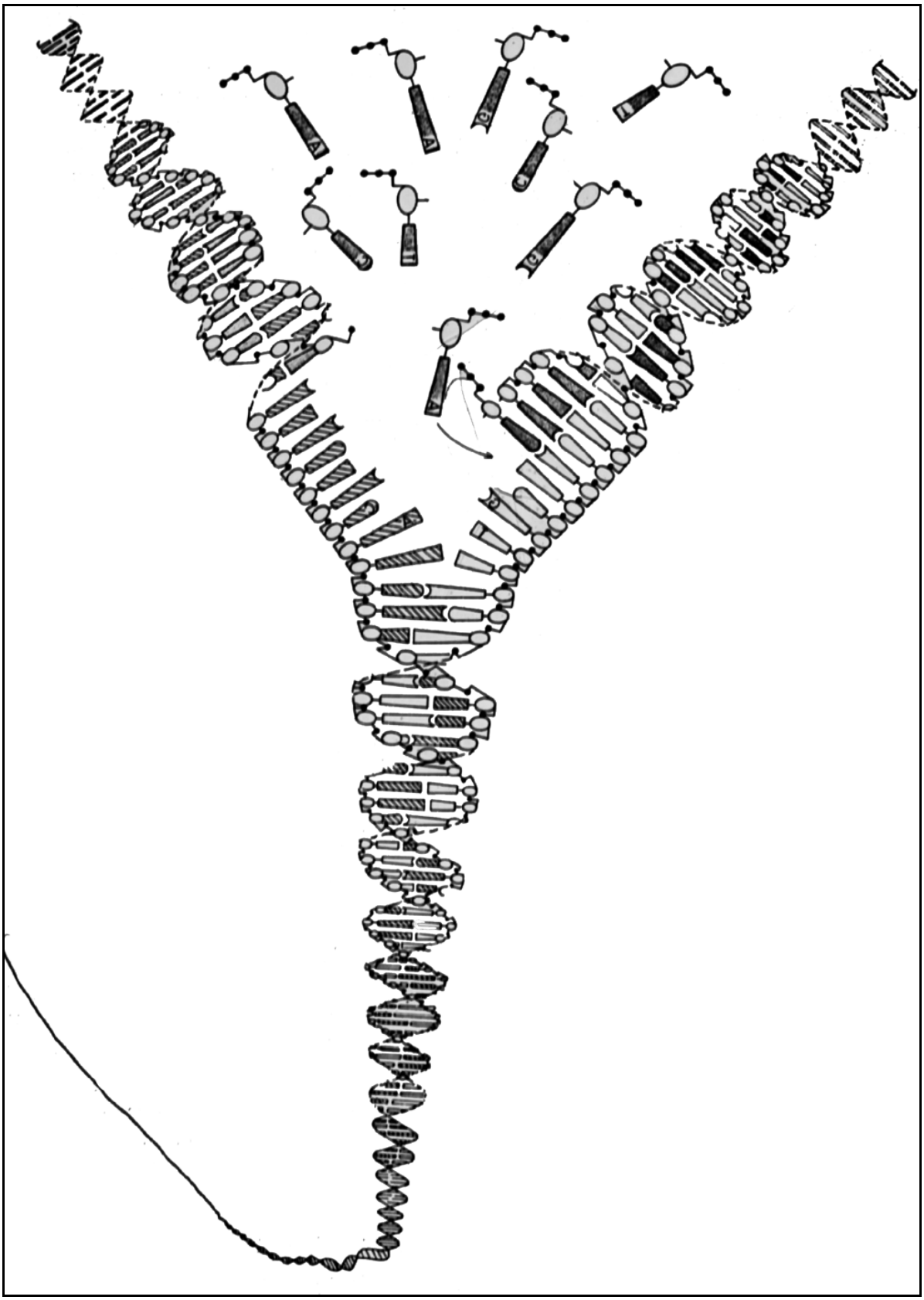


Figure from: D.P. Clark and L.D. Russell (1997) *Molecular Biology Made Fun and Simple* with permission from Cache River Press, Vienna, IIllinois, USA.

Each of the separated strands of DNA serves as a template for the synthesis



of a complementary new strand. Individual nucleotides from the nucleotide pool can recognize their partners and align on each template strand by forming hydrogen bonds, but this method alone would be inefficient. The enzyme DNA polymerase III (pol III) efficiently joins correctly matched nucleotide pairs.

Polarity

The newly synthesized nucleic acid strand always occurs in the direction of 5' to 3'. This is based upon the fact that the 5' phosphate group of the new nucleotide joins to the 3' positions of the nucleotide previous nucleotide.

The hydrogen bonding to the template strand and stringing together of the incorporated nucleotides in the newly synthesized strand occurs at a replication fork. The DNA polymerase catalyzes the elongation of the newly formed strand. In fact, a primary function of DNA polymerase is to

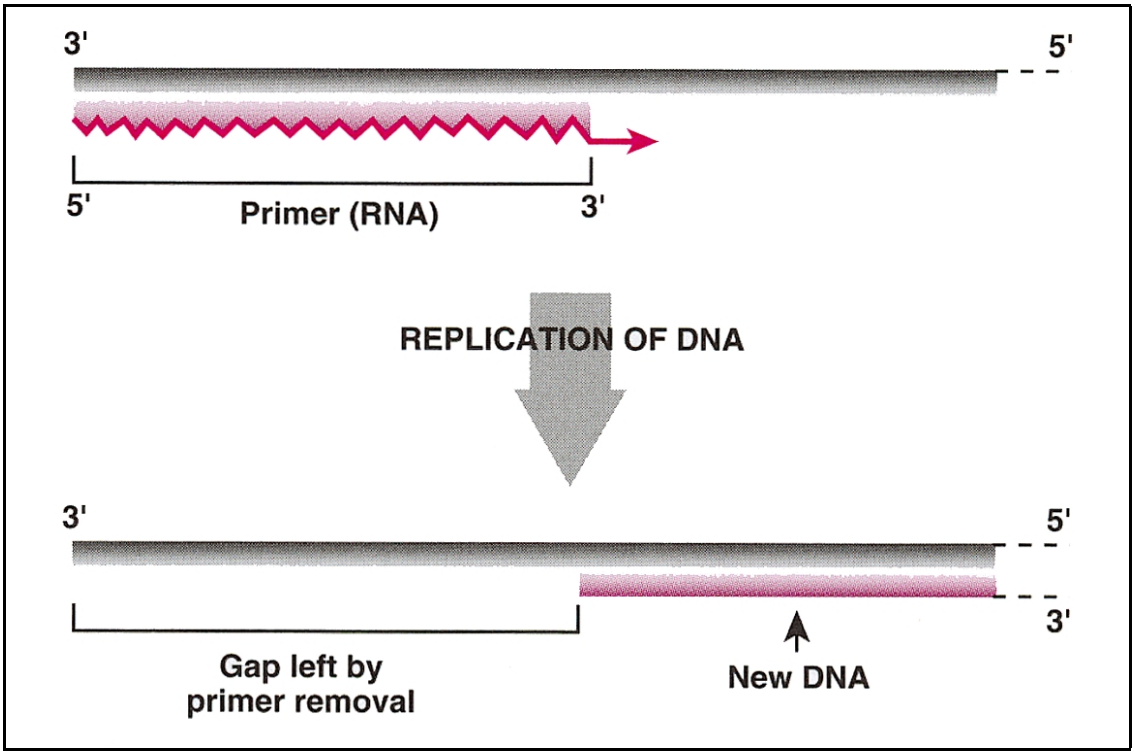
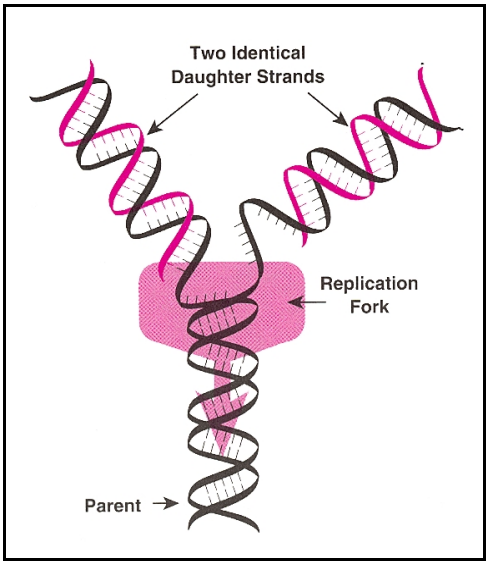
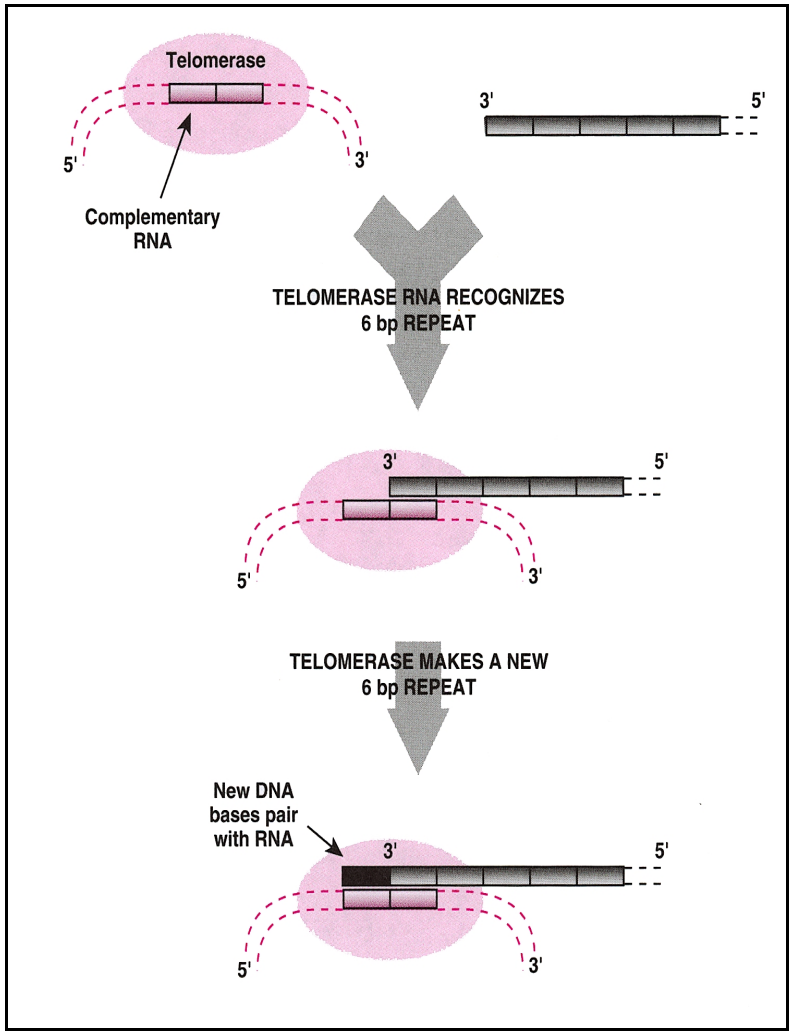


Figure from: D.P. Clark and L.D. Russell (1997) *Molecular Biology Made Fun and Simple* with permission from Cache River Press, Vienna, IIllinois, USA.

elongate the new strand. The DNA polymerase by itself cannot actually start the new strand. What is actually needed to start the process is a

primer and this is actually an RNA. A short segment of primer made up of RNA binds to the separated strands of DNA and primes or starts the synthesis of the new DNA strand at the end of the chromosome being duplicated. This RNA primer is later removed leaving a gap where it was originally located. Thus, during each replication cycle the chromosome is shortened due to the loss of the RNA primer. Once past the replication fork, the newly formed DNA strand forms a double helix structure with the template strand.





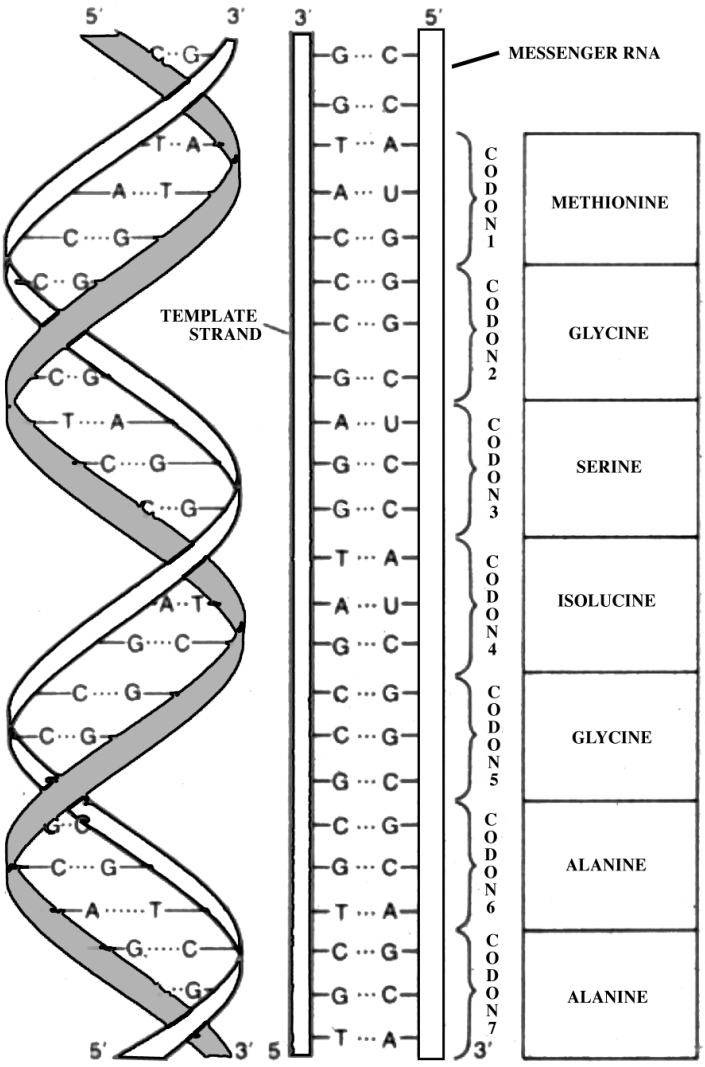
Figures from: D.P. Clark and L.D. Russell (1997) *Molecular Biology Made Fun and Simple* with permission from Cache River Press, Vienna, IIllinois, USA.

This problem is dealt with by an enzyme called **telomerase** that acts at the telometric end of the chromosome. Telomeres consist of six base pairs repeated about 2,000 times. Telomerase contains a small amount of RNA complementary to this six base pair repeat, recognizes the telomere, and synthesizes new nucleotide sequences to fill in the gap left by the RNA primer that was responsible for initiating the DNA replication.

GENE EXPRESSION

Overview

The term gene expression often refers to the production of mRNA from the DNA template. However, in order for the expressed gene to actually result in a functional or phenotypic change, the mRNA must be translated into protein. The overall process is depicted to the left. The process of converting DNA to mRNA is called transcription. The process of converting mRNA into protein is called translation.



Transcription = the generation of single stranded RNA that is identical in sequence with the coding strand of DNA.

Transcription unit = a sequence of DNA that can be transcribed by RNA polymerase into a single RNA beginning at a startpoint (**initiation**) and end at a **terminator**.

Upstream sequences are 5' (prior) to the startpoint of transcription, **downstream** sequences are 3' (after) a particular sequence.

In the process of transcription, related genes often are transcribed simultaneously.

Gene family = a set of genes whose exons are related

> Gene families are believed to have arisen by duplication of an ancestral gene with different members of the family diverging as a consequence of mutations occurring during evolution. This may result in evolution of proteins that are optimized for specific functions (e.g., fetal hemoglobin has a higher affinity for oxygen than adult hemoglobin). Alternatively, mutations may lead to inability to produce a functional protein. These latter genes are referred to as **pseudogenes.**

Gene cluster = a group of genes that are adjacent and are identical or related to each other
Housekeeping genes = genes that encode proteins that provide basic functions in all cells and are always expressed (**constitutive**ly expressed); expression level is constant

RNA Polymerases

RNA polymerases are essential in the initiation of transcription. While there are three different RNA polymerases, RNA polymerase II is the type involved in gene expression that ultimately results in production of protein.

RNA polymerase I -- located in the nucleolus; transcribes genes that encode ribosomal RNA
RNA polymerase II -- located in the nucleoplasm; produces heterogeneous RNA that becomes mRNA after processing and splicing
RNA polymerase III -- located in the nucleoplasm; synthesizes small RNA such as tRNA

There are several steps to consider in understanding gene expression. These are tabulated below as control points for gene expression.

Potential control points for gene expression:
Activation of gene structure
Initiation of transcription
Elongation of transcript

Capping the RNA transcript
Splicing the RNA transcript
Cleaving the RNA transcript
Polyadenylation of the RNA transcript
Transport into the cytoplasm
Susceptibility of RNA to degradation
Translation of mRNA into protein
Post-translational modification of protein
Transport (and secretion) of the protein
Proteolytic cleavage of the protein
Interaction of protein with natural inhibitors

Gene activation = the state of DNA that allows it to be transcribed
• Structural changes in chromatin - unwinding of DNA strands
• Transcribable DNA has increased sensitivity to DNase
• Transcribable DNA is often undermethylated
CpG-rich islands consisting of mostly unmethylated CpG are often found upstream of consitiutively transcribed genes

Transcription

RNA polymerases are necessary but not sufficient for transcription. RNA polymerases bind to the **promoter** region of DNA upstream of the gene that will be transcribed. Accessory factors, called **transcription factors**, are also required.

Promoter = region of DNA where binding of RNA polymerase occurs to initiate transcription
Promoters are **cis-acting** sites on DNA

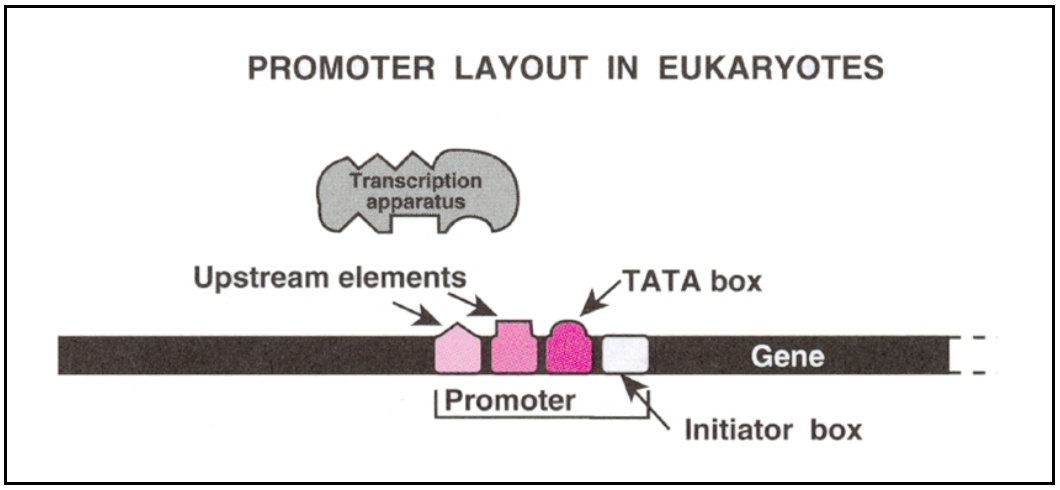


Figure from: D.P. Clark and L.D. Russell (1997) *Molecular Biology Made Fun and Simple* with permission from Cache River Press, Vienna, Illlinois, USA.

Promoters contain short conserved sequences recognized by RNA polymerase and different transcription factors

Hence, transcription factors are **tran-acting** factors that recognize certain cis-acting sites in the promoter

The complex of transcription factors that bind to specific promoter sequences in a particular gene are believed to control the initiation of transcription.

Promoters are generally greater than 100 base pairs, contain several conserved DNA sequences, are upstream but relatively close to the startpoint of initiation, and are protected from nuclease digestion by the binding of trans-acting factors.

Short **consensus sequences** within promoters are called **boxes** and function to help position polymerase correctly help determine the strength of the promoter.

TATA box (consensus sequence = TATAAAA)
CAAT box (consensus sequence = GGCCAATCT)
GC box (consensus sequence = GGGCGG)

Enhancers

First described as 72 base pair repeat elements in SV40
Enhance initiation

May be located relatively far from the startpoint
May be upstream or downstream from the startpoint
May be bound by transcription factors
May be in either orientation
May reside within introns
Consensus sequences are contiguous over short regions
Have a dense concentration of protein binding sites
Even when moved far away from a gene, enhancers can still enhance transcription

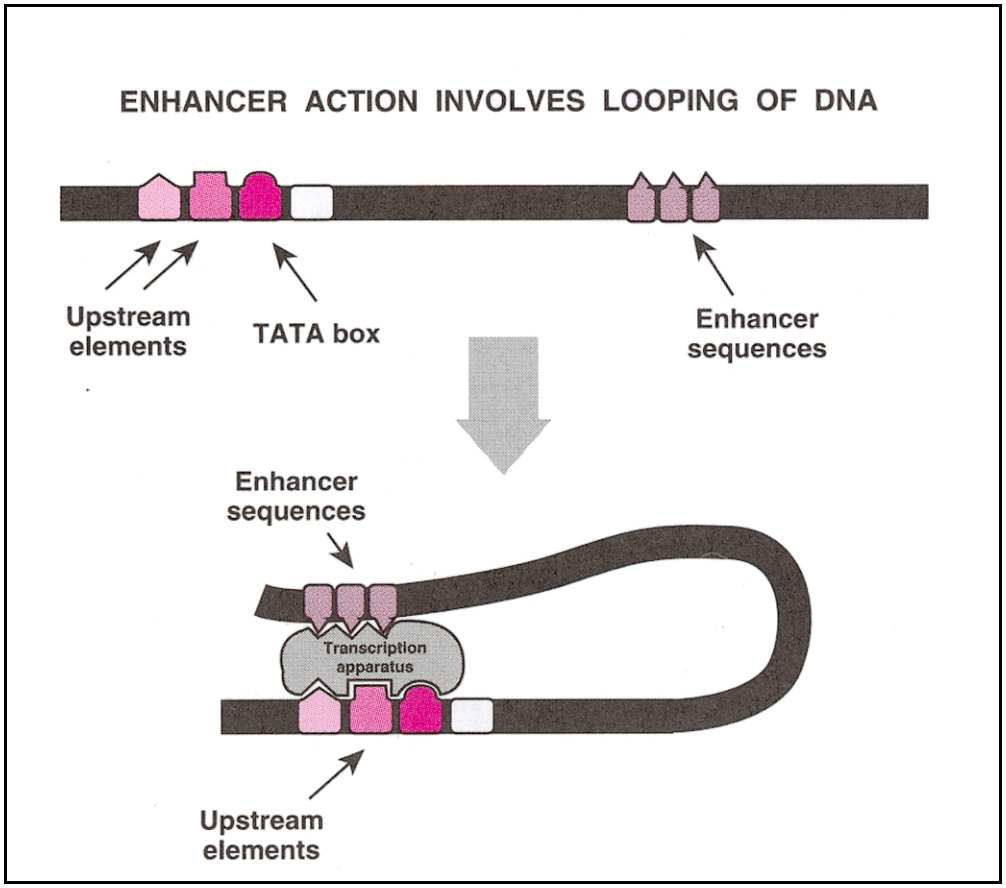


Figure from: D.P. Clark and L.D. Russell (1997) *Molecular Biology Made Fun and Simple* with permission from Cache River Press, Vienna, IIllinois, USA.

Response Elements

The consensus sequences that uniquely identify groups of genes
These groups of genes are under common control
May share consensus sequences recognized by specific transcription factors
Transcription factors that recognize a response element coordinate the transcription of all the genes that have that response element

Are found within promoters and enhancers

Heat shock response element (HSE) - These response elements are localized in the promoter region of heat shock genes and are turned on in response to heat shock.

Glucocorticoid response element (GRE) - These response elements govern the response to steroid hormones and are typically located about 250 bp upstream from the startpoint of transcription.

Metal response element (MRE) enhance transcription of genes in response to heavy metals. An example of a gene with an MRE is the metallothionein gene.

TPA (phorbol ester) response element (TRE). TPA is a tumor promoting agent classically used in skin painting carcinogenicity studies. In response to TPA, AP1 transcription factors such as Jun and Fos bind to the TRE.

Transcription factors have DNA binding motifs

Many DNA-binding proteins can be grouped into classes based upon their use of structural motifs for recognition

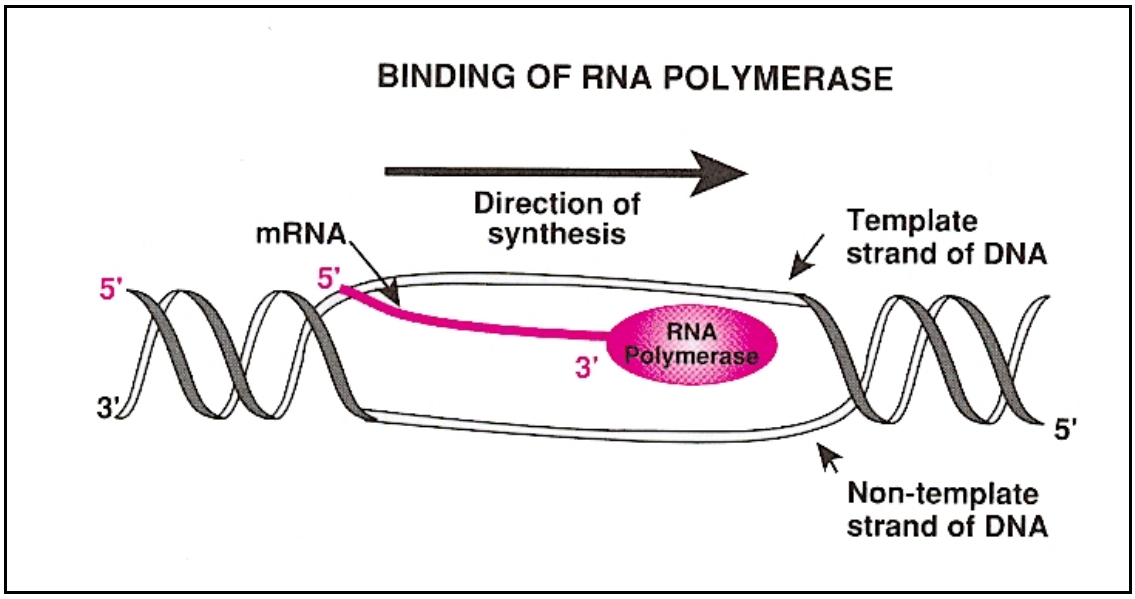
Helix-turn-helix structures - a structural motif that provides specificity of target recognition

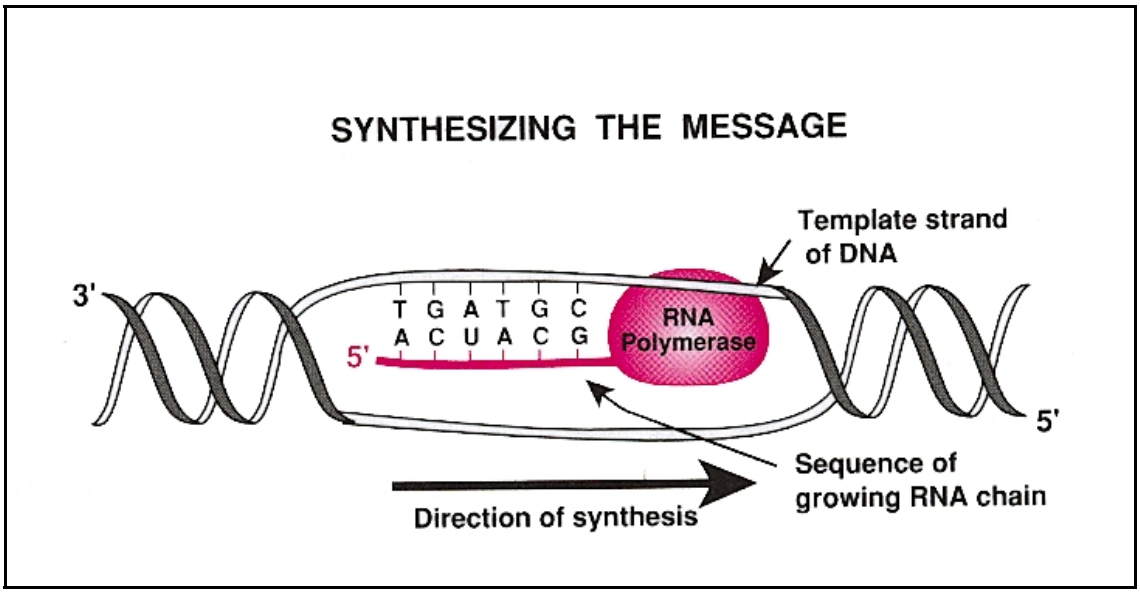
Zinc fingers - structural motif that has cysteine and histidine residues that bind zinc

Steroid receptors - regulatory proteins that include receptors for steroid hormones, retinoids, vitamin D and thyroid hormones

Leucine zippers - leucine-rich stretches of amino acids that play an important role in differentiation and development

Important in dimer formation typical of **AP1 binding proteins,** especially **homodimers** (Jun-Jun) and **heterodimers** (Jun-Fos)





Figures from: D.P. Clark and L.D. Russell (1997) *Molecular Biology Made Fun and Simple* with permission from Cache River Press, Vienna, IIllinois, USA.

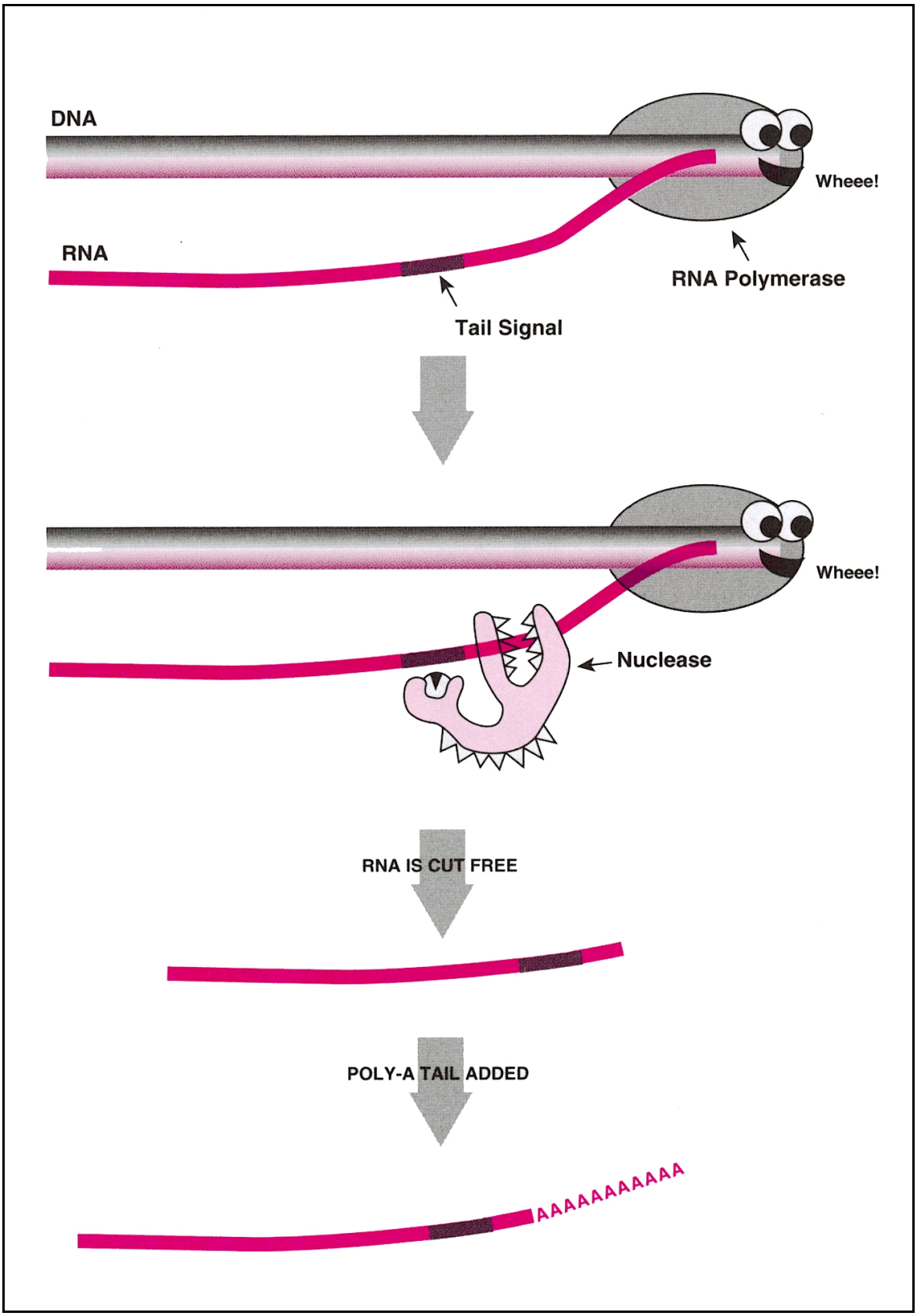


Figure from: D.P. Clark and L.D. Russell (1997) *Molecular Biology Made Fun and Simple* with permission from Cache River Press, Vienna, IIllinois, USA.

Just as initiation of mRNA synthesis is highly regulated, the termination of mRNA synthesis is regulated by the presence of specific terminator sequences in the gene being transcribed. The signal is two inverted repeats followed by a string of A's. As a consequence of this terminator sequence,

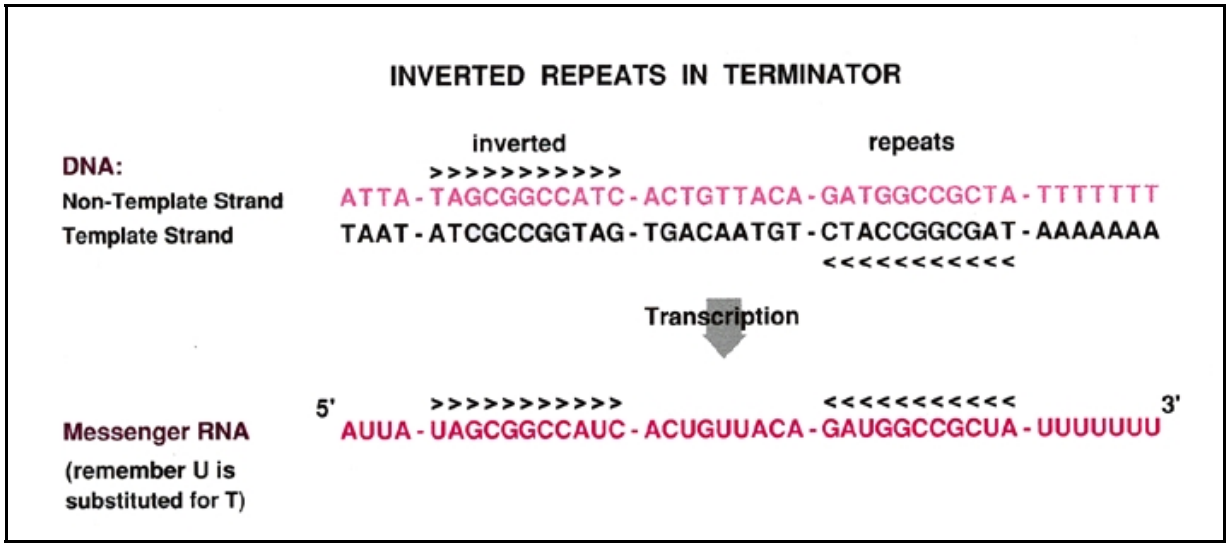


Figure from: D.P. Clark and L.D. Russell (1997) *Molecular Biology Made Fun and Simple* with permission from Cache River Press, Vienna, IIllinois, USA.

the mRNA contains complementary inverted repeats which form a stem and loop or hairpin structure and the string of A's in the DNA template leads to a string of U's in the mRNA.

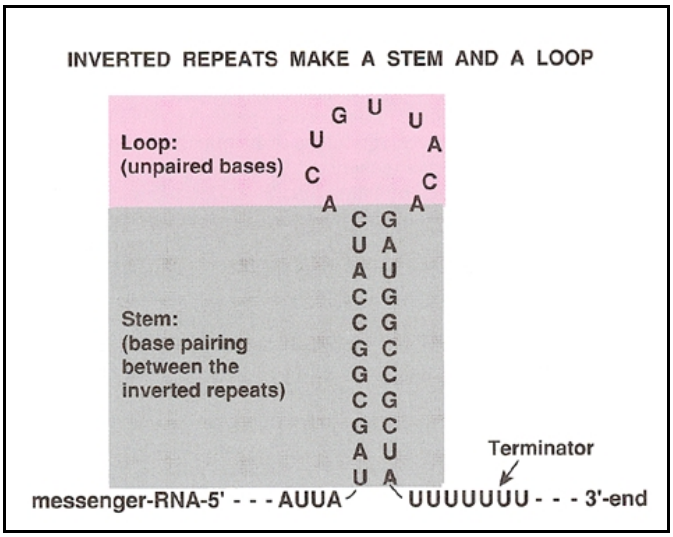


Figure from: D.P. Clark and L.D. Russell (1997) *Molecular Biology Made Fun and Simple* with permission from Cache River Press, Vienna, IIllinois, USA.

What basically happens is that when the RNA polymerase reaches the hairpin structure, it slows down or stops momentarily. Since the string of U's complementary to the A's forms a weak structure, the RNA and DNA fall apart, the RNA polymerase falls off, and transcription stops.

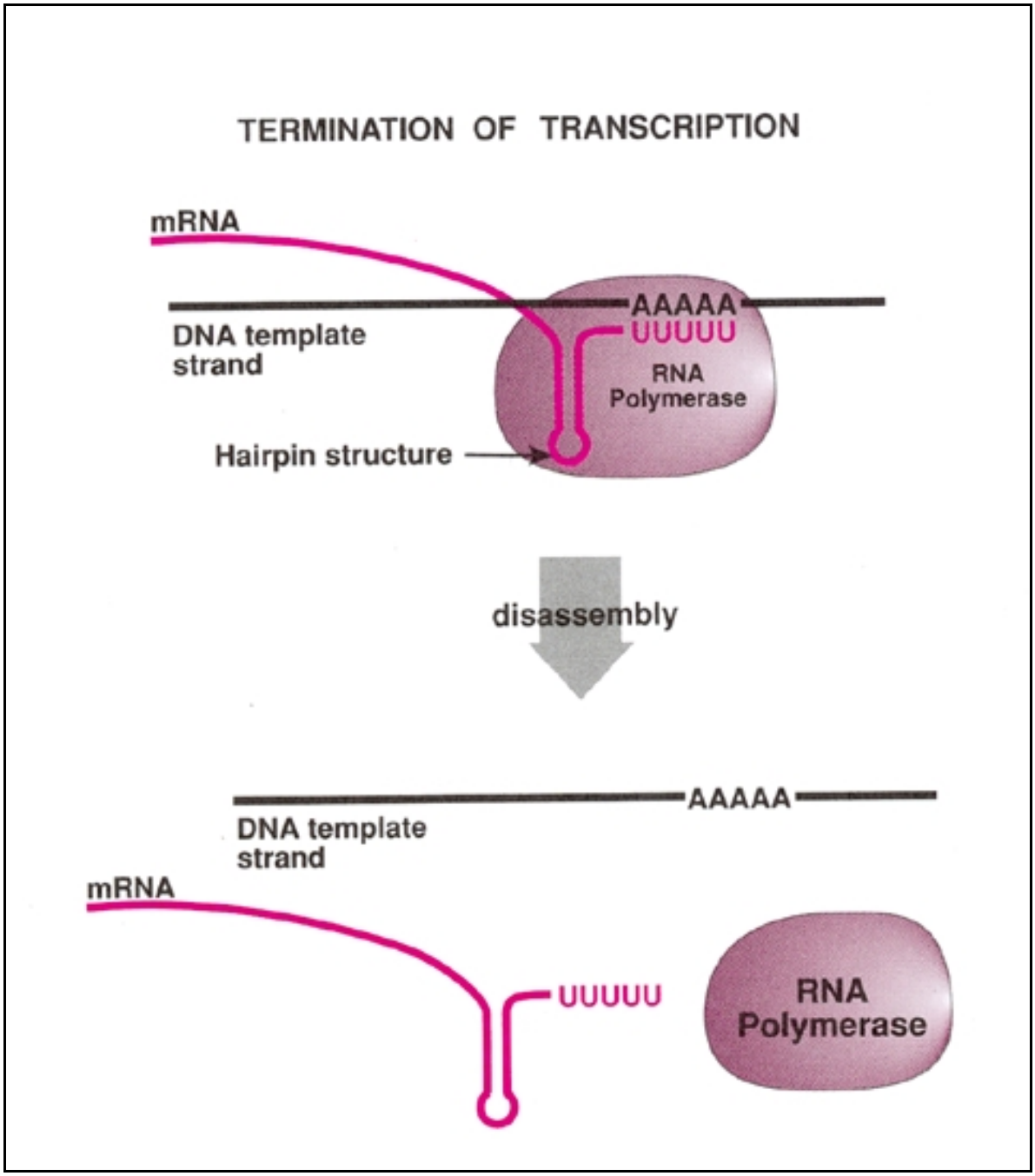


Figure from: D.P. Clark and L.D. Russell (1997) *Molecular Biology Made Fun and Simple* with permission from Cache River Press, Vienna, IIllinois, USA.

Splicing

Splicing is the procedure in which introns which separate the coding sequences (exons) are removed in preparation for translation of the mRNA into protein. Splicing is complex and allow for creation of different proteins from the same mRNAs. Alternative splicing allows for the formation of more than one protein from a single gene.

This can occur by initiating transcription at different promoter sequences that are upstream of the gene coding sequences or by adding the polyA tail at alternative sites downstream of the gene coding sequence. Splicing can also involve exon cassette selection or trans-splicing. Trans-splicing results in mRNA derived from two separate original RNA molecules.

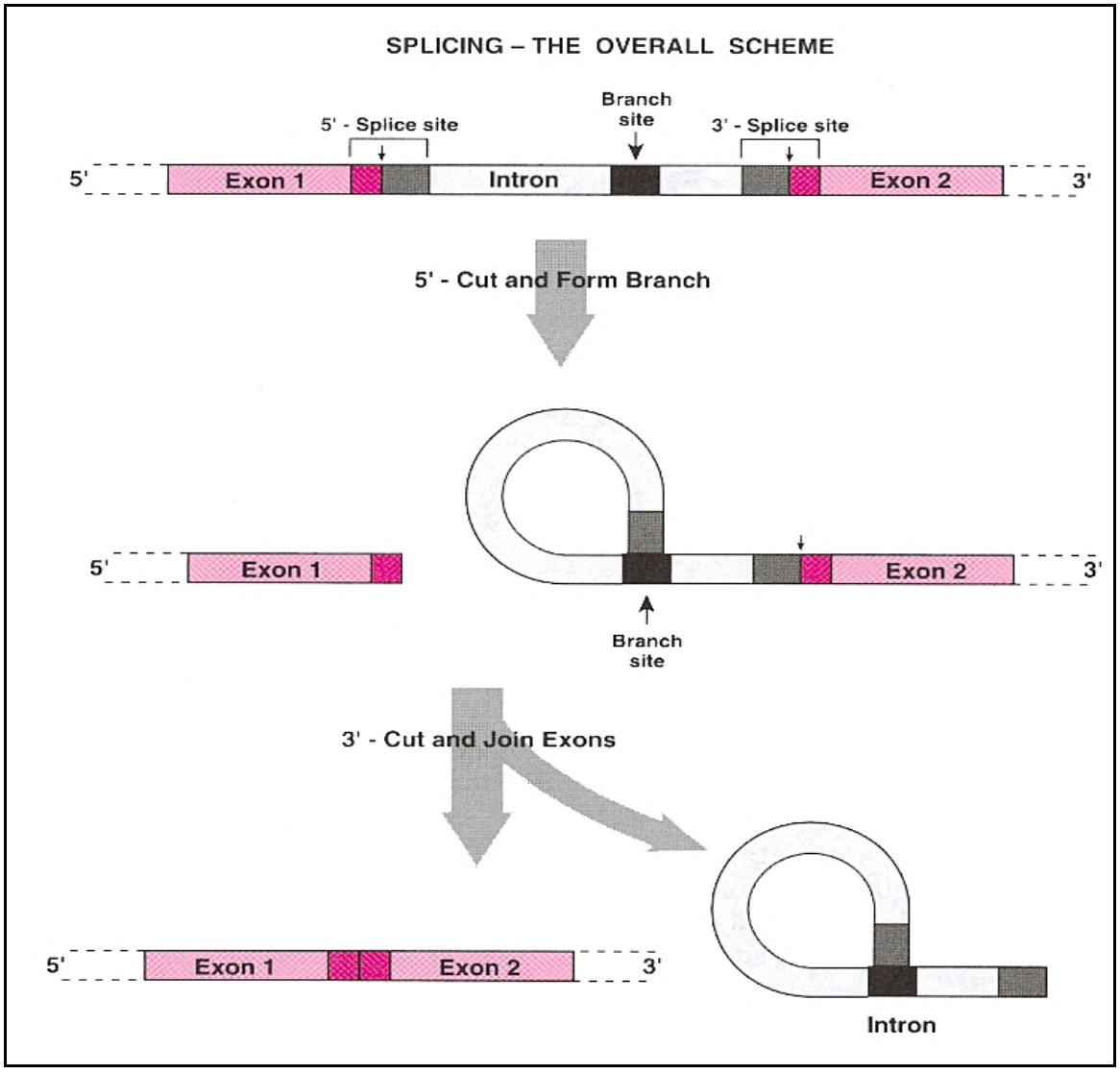


Figure from: D.P. Clark and L.D. Russell (1997) *Molecular Biology Made Fun and Simple* with permission from Cache River Press, Vienna, IIllinois, USA.

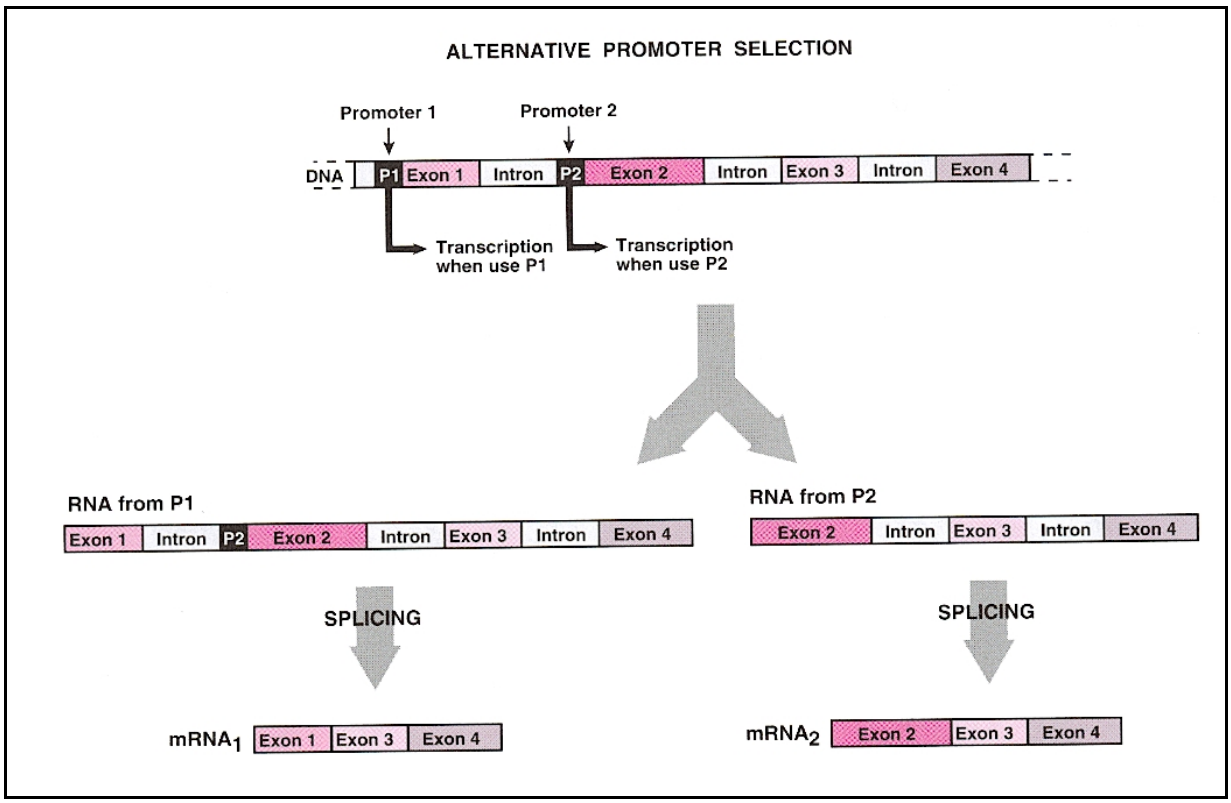


Figure from: D.P. Clark and L.D. Russell (1997) *Molecular Biology Made Fun and Simple* with permission from Cache River Press, Vienna, IIllinois, USA.

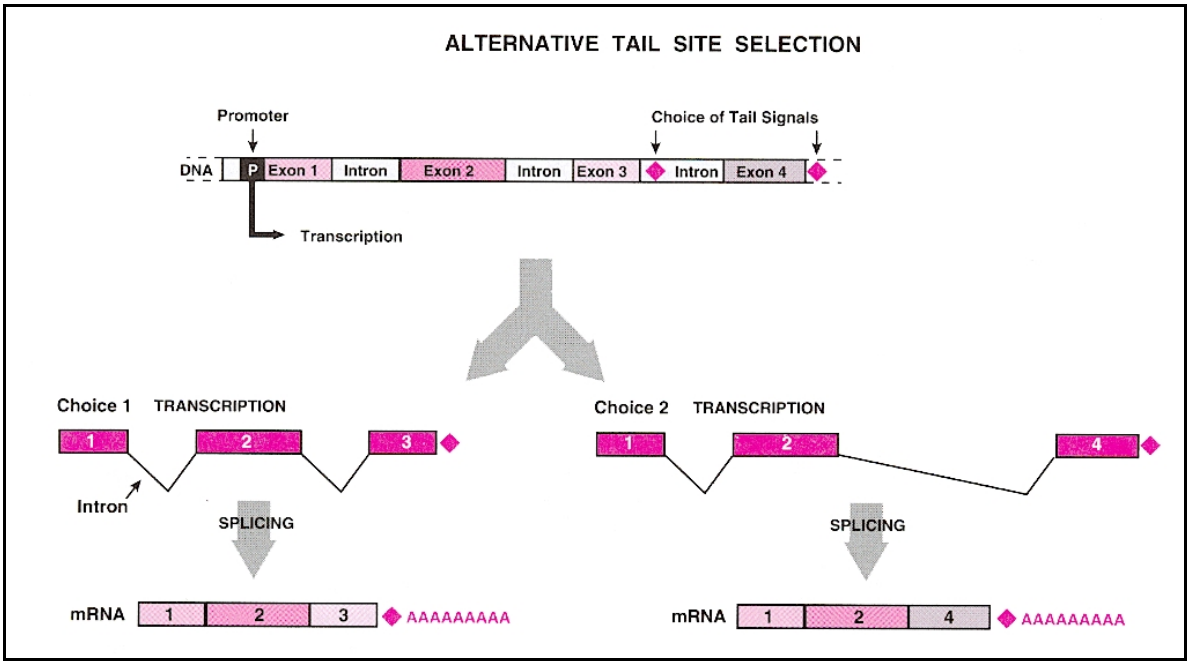


Figure from: D.P. Clark and L.D. Russell (1997) *Molecular Biology Made Fun and Simple* with permission from Cache River Press, Vienna, IIllinois, USA.

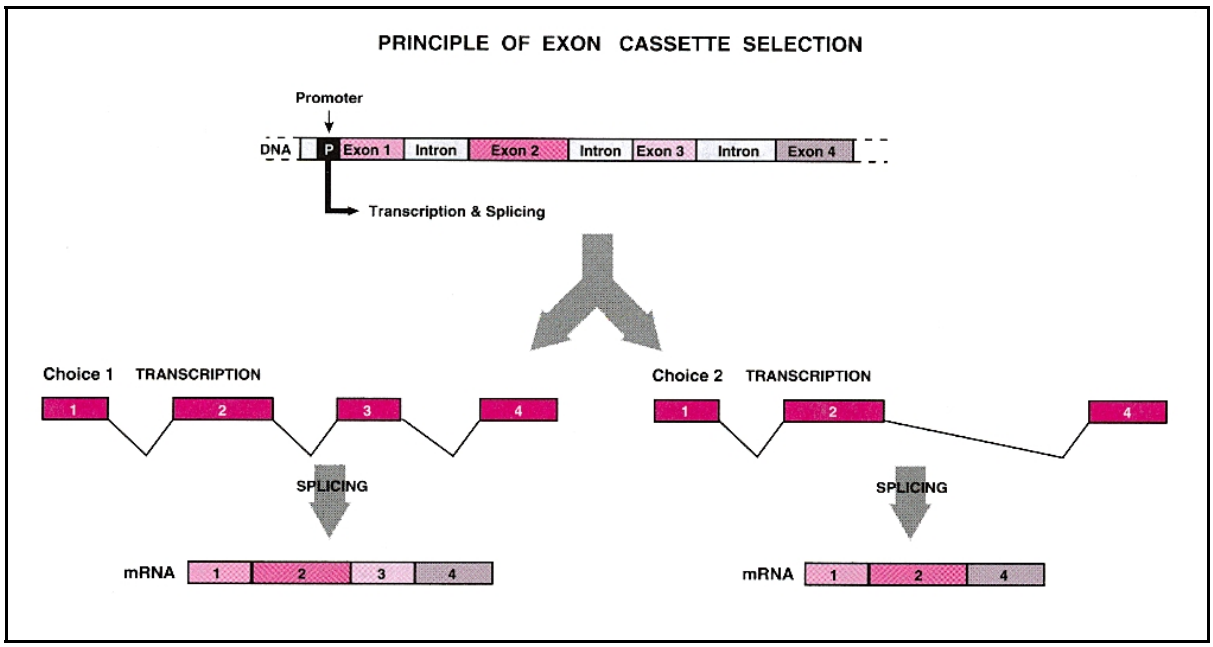


Figure from: D.P. Clark and L.D. Russell (1997) *Molecular Biology Made Fun and Simple* with permission from Cache River Press, Vienna, IIllinois, USA.

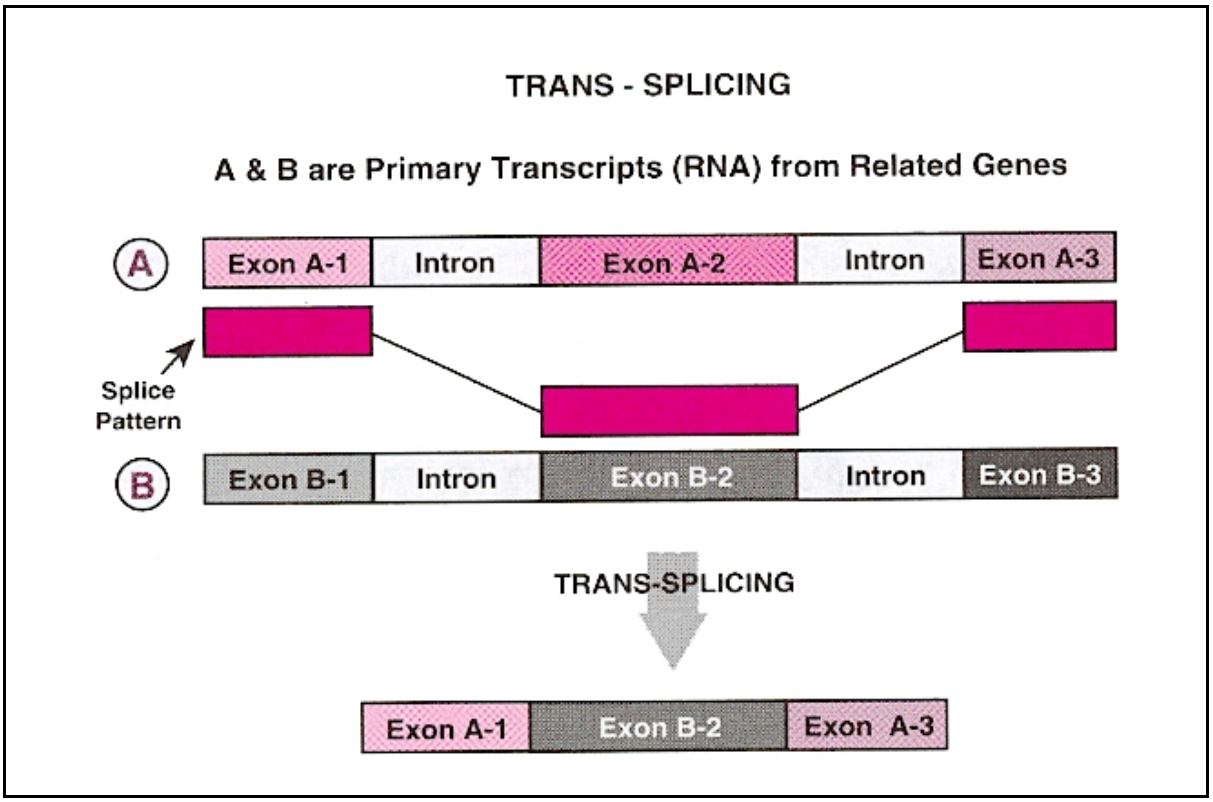


Figure from: D.P. Clark and L.D. Russell (1997) *Molecular Biology Made Fun and Simple* with permission from Cache River Press, Vienna, IIllinois, USA.

Caps and Polyadenylation Tails

Prior to its transport from the nucleus to the cytoplasm, mRNA is capped at its 5' end and a polyA tail is added at its 3' end. The cap consists of a guanine that is added to the first nucleotide in reverse orientation via a triphosphate link. The 5' terminal guanine (G) is often methylated.

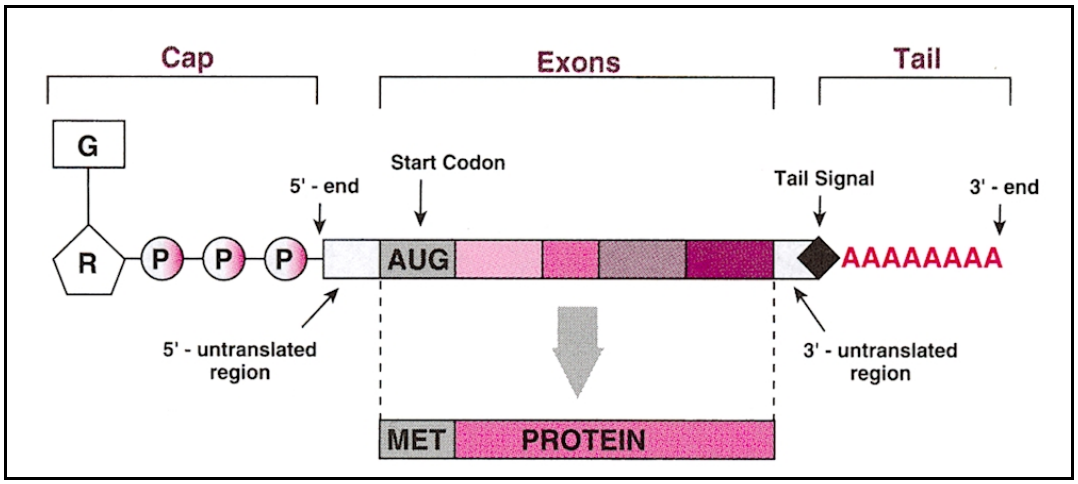


Figure from: D.P. Clark and L.D. Russell (1997) *Molecular Biology Made Fun and Simple* with permission from Cache River Press, Vienna, IIllinois, USA.

The cap with its complex methylated structure is believed to be important for maintaining the stability of the mRNA and is essential for translation.

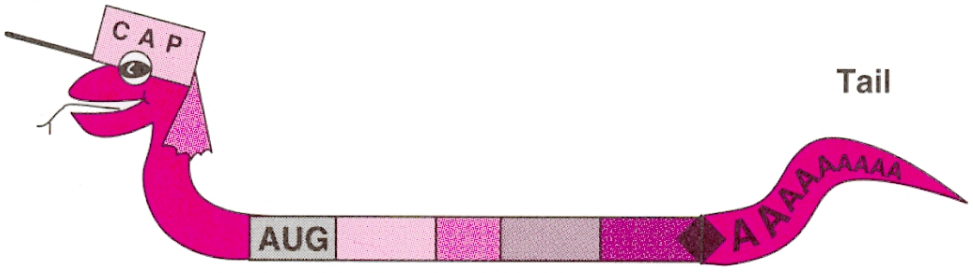


Figure from: D.P. Clark and L.D. Russell (1997) *Molecular Biology Made Fun and Simple* with permission from Cache River Press, Vienna, IIllinois, USA.

There is a recognition sequence at the 3' end of the mRNA molecule consisting of AAUAAA. While the RNA polymerase that is fabricating the mRNA from the DNA template continues beyond this AAUAAA recognition sequence (tail signal), a specific endonuclease recognizes this sequence and cuts the mRNA molecule at a point 10 to 30 bases downstream from the AAUAAA sequence. The enzyme poly(A) polymerase then adds a string of

100 to 300 adenine residues to form the polyA tail. The polyA tail provides the molecular biologist with a handy way to recognize the 3' end of mRNA. In terms of the biological significance of the polyA tail, little is known. It has been hypothesized that the polyA tail stabilizes the mRNA, and it has also been suggested that the polyA tail is involved in the export of mRNA from the nucleus.

Once mRNA has been capped and tailed and undergone splicing, it can then move through the nuclear pores and into the cytoplasm.

Translation of mRNA into Protein

Protein synthesis occurs in association with the ribosomes and involves several complex procedures involving rRNA and various tRNAs and accessory factors. The sequence of amino acids that constitute the protein product are derived from decoding the genetic code contained in the codons (see Genetic Code Table p. 2-7). Codons are groups of three RNA or DNA bases which encode a single amino acid. tRNAs recognize and bind to a codon on the mRNA and carry the amino acids to a ribosome. The mRNA bases are read in groups of three starting from the 5' end and always beginning with the start codon, AUG. There can be more than one start codon sequence at the 5' end of mRNA. For example, three AUG sequences are depicted below:

5' GAUAUGUAUGCGAUGCCGGAAACAUCUAAGGA 3'

Depending upon where translation starts (which start codon is used), entirely different products can result as the subsequent codons are translated in successive groups of three into specific amino acids. The three possibilities in the above example are referred to as reading frames. A sequence of mRNA that begins with a start codon and ends with a stop codon (a nonsense codon) and can thus be translated into protein is referred to as an open reading frame. Termination (stop, nonsense) codons include UAA, UAG, and UGA. All proteins start with a methionine. The codon AUG specifies methionine and is sometimes called the initiator codon. A variety of control mechanisms and the possibility of perturbations in translational control factors can undoubtedly account for cellular abnormalities. This is an area of active investigation. The stop codons signal for the termination of translation. In the absence of a stop codon (e.g., defective mRNA) the entire process may become stalled without production of a useful protein.

mRNA Degradation

Once mRNA is transcribed and translated it is subject to degradation by a regulated process. Half lives of mRNAs can vary from minutes to days. Most mRNAs need to be translated in order for there to be degradation.

Decapping enzymes, which are known to occur, could destabilize mRNA. Deadenylation of the polyA tail may be a prerequisite for mRNA degradation. Consensus sequences such as AUUUA at the 3' untranslated regions of mRNA in conjunction with 20S protein complexes have been suggested to alter cellular metabolism and precipitate degradation of mRNA.

Summary of Gene Expression

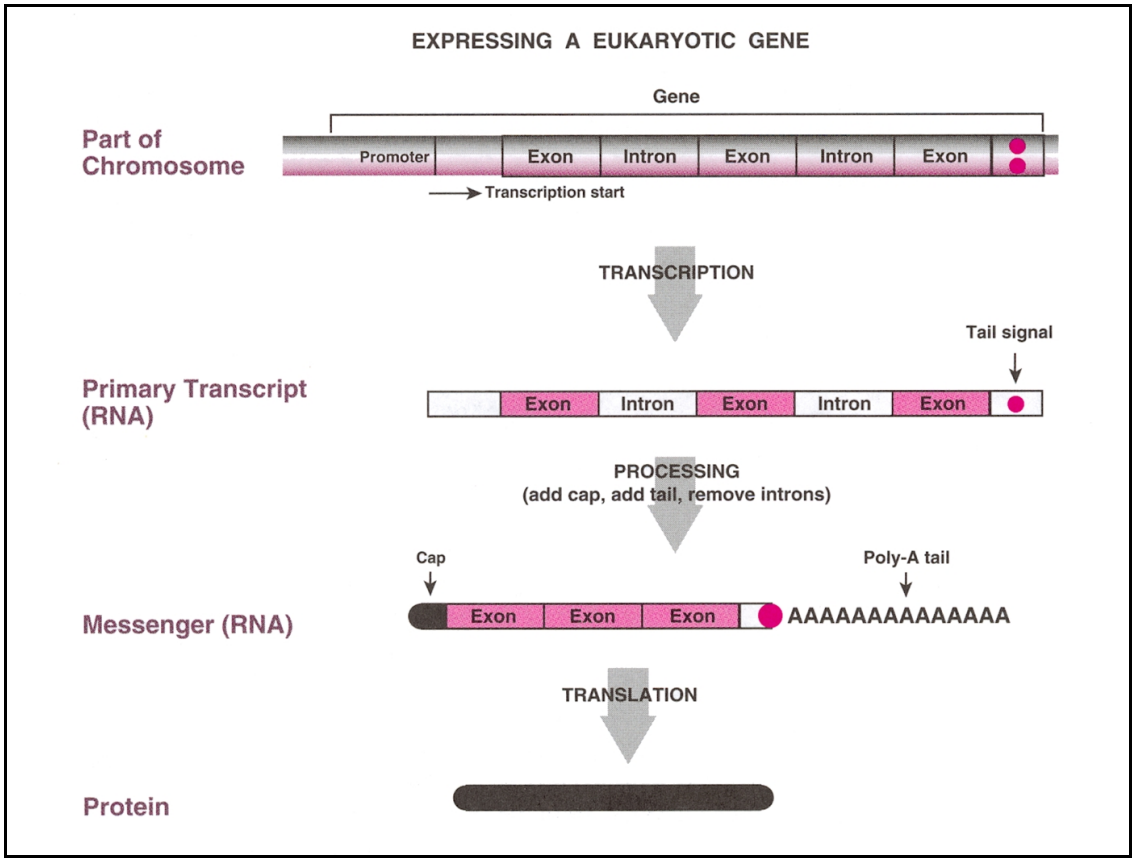


Figure from: D.P. Clark and L.D. Russell (1997) *Molecular Biology Made Fun and Simple* with permission from Cache River Press, Vienna, IIllinois, USA.

RESTRICTION ENZYMES

Synthesis of nucleic acids is catalyzed by specific enzymes. DNA polymerase synthesizes DNA and RNA polymerase synthesize RNA. Degradation of nucleic acids also occurs under the influence of specific enzymes. There are exonucleases and endonucleases. Exonucleases remove residues one at a time from the end of the molecule, while endonucleases cut individual bonds within the molecule. Several endonucleases, derived from bacteria, are used by the molecular biologist to fragment nucleic acids at defined points. These are frequently referred to as restriction enzymes.

Carefully isolated DNA is often too large in length for us to easily work with it; consequently, we need to cut it up into pieces of a manageable size. This is typically done using **restriction enzymes** which bind to DNA at specific recognition sites (specific sequences of bases). The recognition site is typically 4 to 8 bases long and in the form of an **inverted repeat**.

5' -- G - G - A - T - C -C --3'

3' -- C - C - T - A - G - G --5'

Recognition site for BamH1

5' -- G - A - A - T - T - C --3'

3' -- C - T - T - A - A - G -- 5'

Recognition site for EcoR1

5' -- G - T - T - A - A - C -- 3'

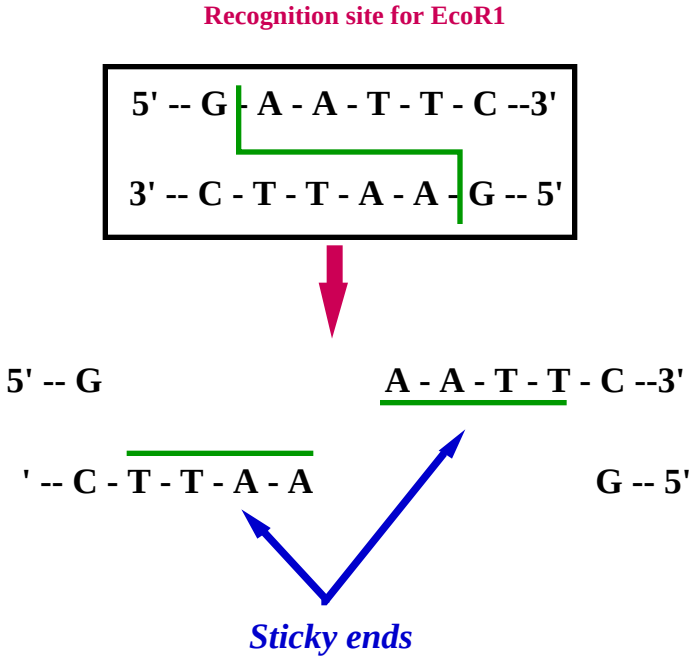
3' -- C - A - A - T - T - G --5'

Recognition site for Hpa1

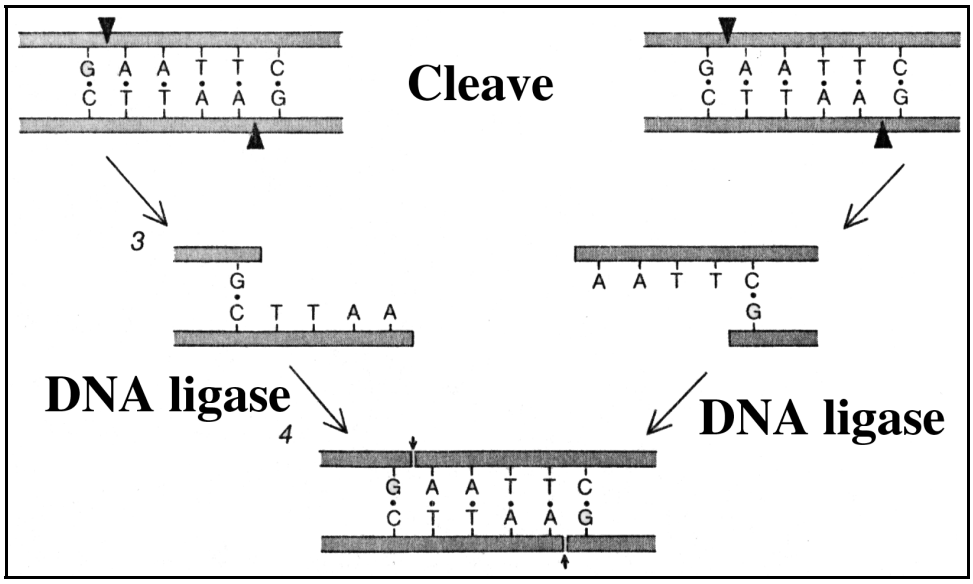
Restriction enzymes are bacterial endonucleases and they are widely used in the molecular biology laboratory. Like many other situations in molecular biology, the nomenclature of restriction endonucleases is historical and may not have apparent logic.
Restriction enzymes get their name from the bacteria from which they were isolated in combination with some input from the molecular biologist. The

restriction enzyme **EcoRI** originated from **E**schericia **co**li strain **R**Y13 and the "**I**" derives from the fact that this was the first restriction enzyme found in this species of bacteria.

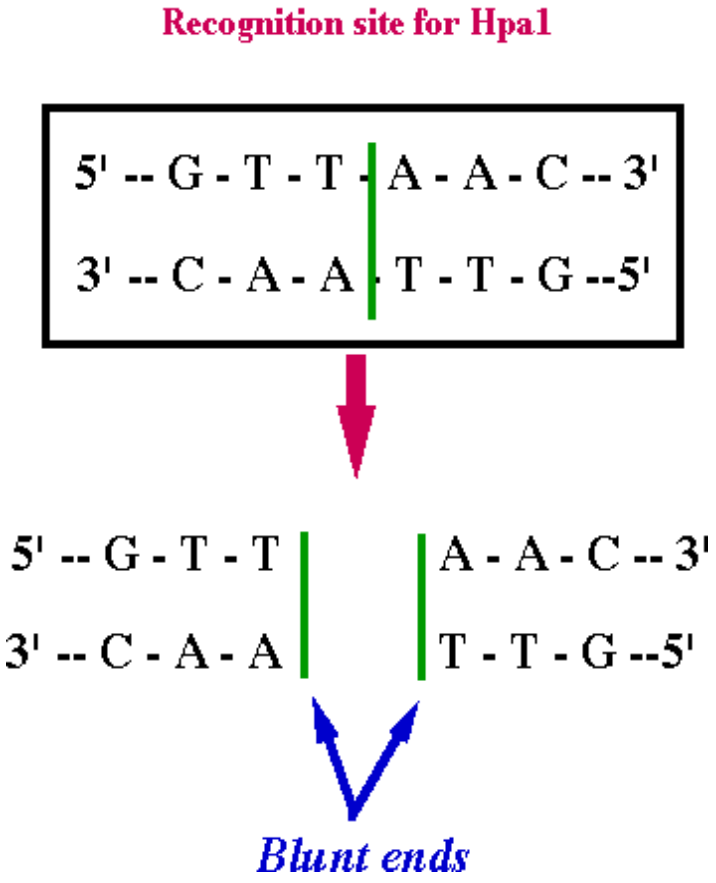
Type II restriction enzymes cut the DNA at the **recognition site** in an known position and yield either blunt ends or sticky ends.



In the presence of **DNA ligase**, the sticky ends can hybridize with complementary nucleotide sequences.



Ligation of blunt ends is inefficient and requires large amounts of DNA ligase while sticky ends are more readily susceptible to ligation.



Restriction Fragment Length Polymorphisms

Restriction fragment length polymorphisms **(RFLPs)** are found as a consequence of fragmenting the DNA with restriction endonucleases. Since each restriction enzyme recognizes a specific sequence, then a single base change will prevent the site-specific cutting of that DNA.

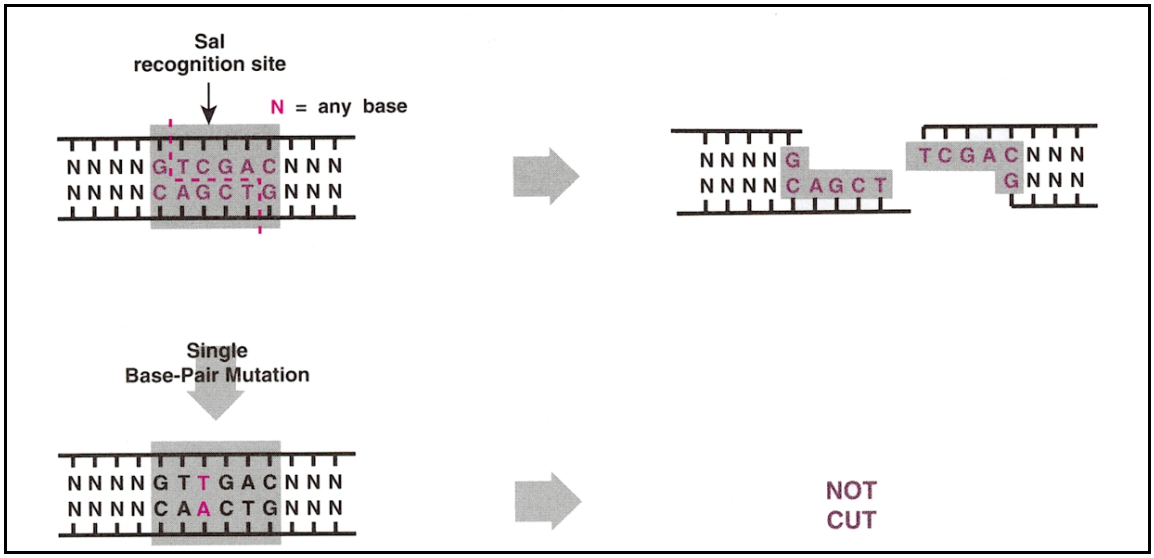


Figure from: D.P. Clark and L.D. Russell (1997) *Molecular Biology Made Fun and Simple* with permission from Cache River Press, Vienna, IIllinois, USA.

Small differences in DNA sequences frequently exist within the genome usually as a result of differences in inherited maternal and paternal alleles. In some instances the difference resides within the recognition site for a specific restriction enzyme. If we cut two related, but different, DNA molecules with the appropriate restriction endonuclease, we may get segments of DNA that differ in length. These can be detected based upon their different size using an electrophoretic gel. The resulting difference between the two fragmented DNA molecules is referred to as an RFLP. Even though we may not appreciate the function of the DNA segments in question, their identification allows use to identify organisms, establish relationships, and detect differences. The RFLP does not need to originate from a coding region of a gene to be a useful marker.

REVERSE TRANSCRIPTION, cDNA, AND AMPLIFICATION

Reverse Transcription

The process of making a segment of DNA from a segment of RNA is called **reverse transcription** and is mediated by the enzyme reverse transcriptase. In fact, this is the mechanism by which retroviruses convert single stranded RNA inserted into a host to double stranded DNA. Since the process reverses the normal flow of genetic information, it is called reverse transcription.

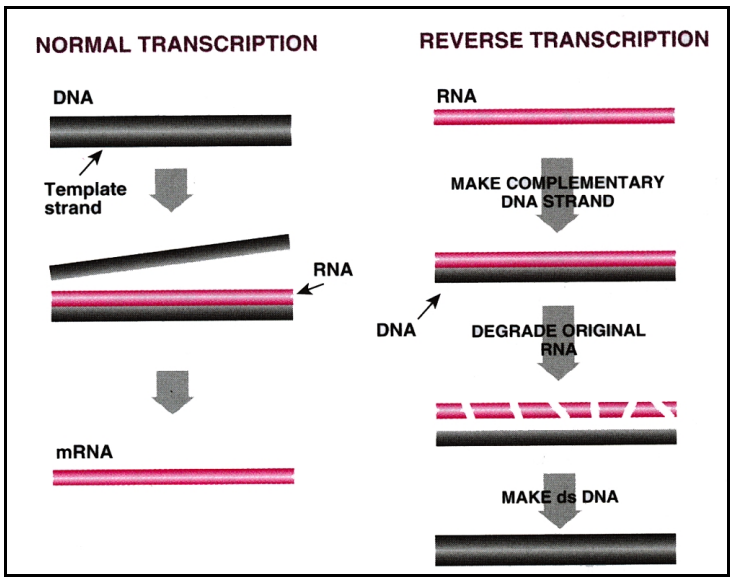


Figure from: D.P. Clark and L.D. Russell (1997) *Molecular Biology Made Fun and Simple* with permission from Cache River Press, Vienna, IIllinois, USA.

cDNA

If we isolate a sample of mRNA from a cell or tissue, we can take advantage of the process of reverse transcription in vitro to make a copy of DNA from

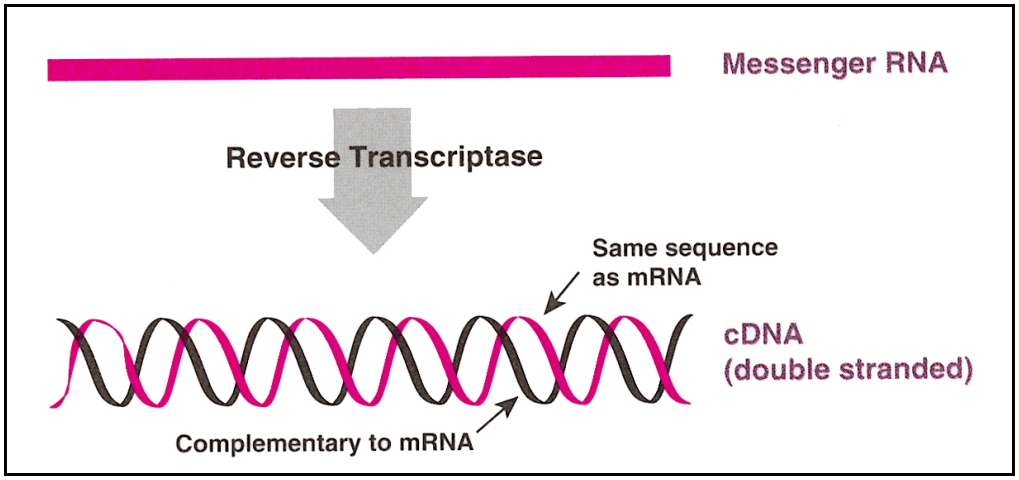


Figure from: D.P. Clark and L.D. Russell (1997) *Molecular Biology Made Fun and Simple* with permission from Cache River Press, Vienna, IIllinois, USA.

the mRNA sample. The newly formed single stranded DNA is complementary, by virtue of the rules of base pairing, to the originally isolated mRNA. This is called complementary DNA (**cDNA**). Double stranded DNA can be derived from the cDNA by base pairing. Since the single stranded cDNA is complementary to the original mRNA, it can be tagged with an appropriate marker (radioactive or fluorescent) and used as a probe to identify the complementary mRNA in cells of another tissue.

Since the cDNA is derived from mRNA originally isolated from the cytoplasm, that mRNA has already been spliced to remove the introns. Consequently, the derived cDNA represents only the exons or coding region of the gene. The cDNA can be isolated and put into an appropriate plasmid or viral vector for cloning and, thus, can be propagated for a variety of uses. In order to obtain enough cDNA for use in the molecular biology laboratory, it can be amplified by a polymerase chain reaction (**PCR**). A **cDNA library** can thus be created to represent the mRNA present in a given tissue. The Cancer Gene Anatomy Project (CGAP) is based upon this technology of isolating mRNA from specific cancers, producing amplified quantities of cDNA from the isolated mRNA by reverse transcription polymerase chain reaction (**RT-PCR**), and establishing a library of expressed genes in specific cancers.

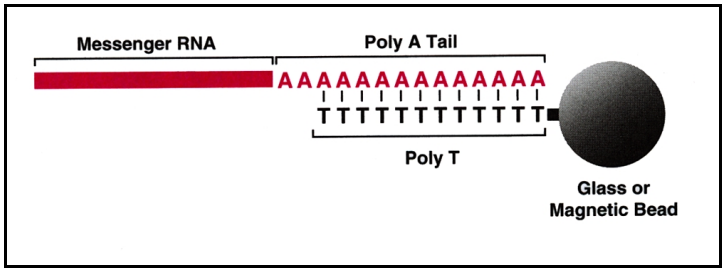


Figure from: D.P. Clark and L.D. Russell (1997) *Molecular Biology Made Fun and Simple* with permission from Cache River Press, Vienna, IIllinois, USA.

Amplification

One of the most powerful in vitro techniques in molecular biology involves amplification of DNA sequences by the **polymerase chain reaction (PCR).** A series of discoveries are important in the application of PCR:

1. Isolation of bacterial **DNA polymerase** and elucidation of its mechanism of action wherein a **primer** is annealed to the template DNA and the DNA polymerase extends the primer by synthesizing new complementary DNA.

2. Isolation of thermostable DNA polymerases.

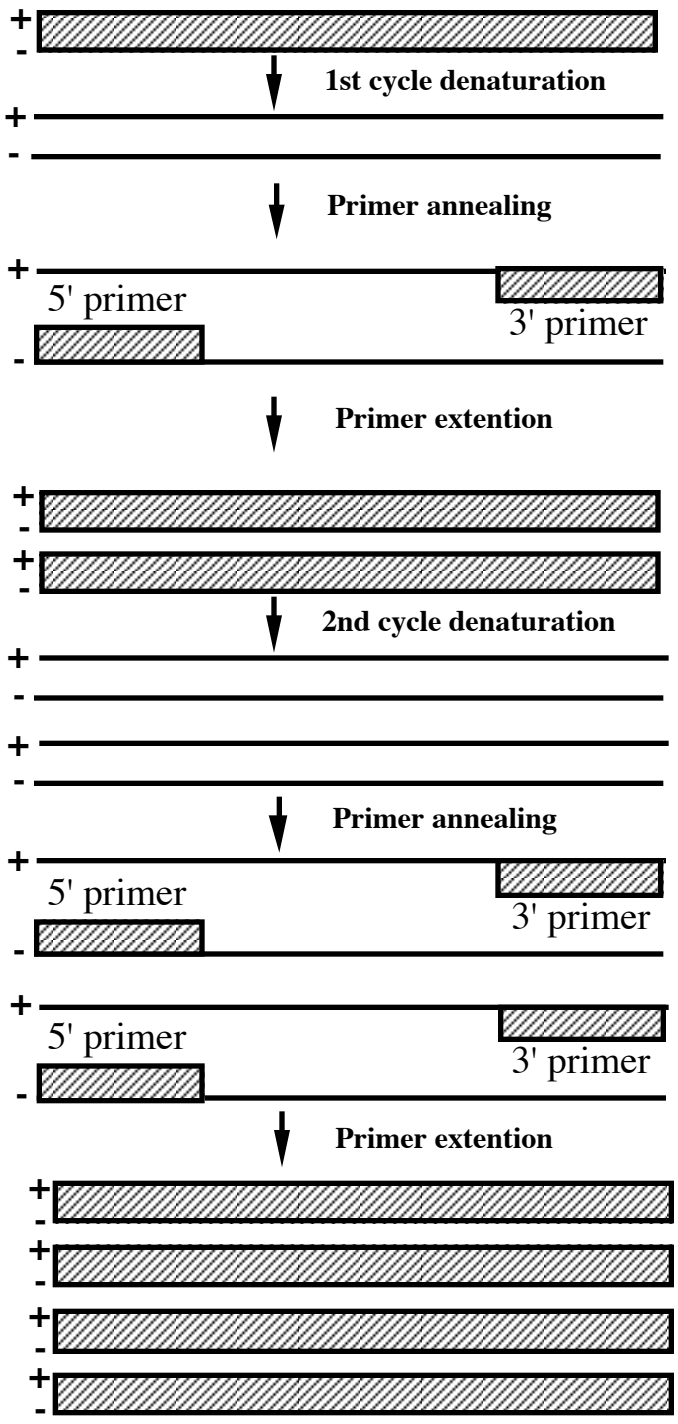
3. Development of a **thermal cycler**.

The reaction requires a template DNA and a primer to function as the anchor for the DNA polymerase. The primer will anneal to the template DNA where it is complementary to the nucleotide string. DNA synthesis then occurs by primer extension. PCR starts with as little as one molecule of double stranded DNA and through successive thermal cycling results in rapid production of millions of copies. After about 20 cycles, about one millions copies are produced from one original molecule. There are actually two primers - one that works at the 5' end and one at the 3' end of the double stranded DNA of interest. Usually the primers are designed to recognize complementary sequences just upstream and downstream of the DNA of interest so that the entire sequence of interest is amplified.

As a practical consideration, it is currently difficult to amplify DNA fragments greater than 4 kb in length. However, the good news is that relatively crude preparations of DNA can be used in the PCR reaction.

The PCR cycles, which as successively repeated produce more and more copies of the DNA segment of interest, consist of three steps: denaturation, annealing of the primer, and primer extension.

Polymerase Chain Reaction



NUCLEIC ACID HYBRIDIZATION

DNA can be denatured and renatured. The process of strand separation is called denaturation (colloquially called melting). Heating or exposure to low salt concentrations destabilizes the noncovalent forces that stabilize the double helix, and this results in strand spearation. DNA denaturation occurs over a narrow temperature range. For a particular DNA molecule, the melting temperature is influenced by the proportion of GC base pairs. The more GC base pairs, the higher the temperature necessary to denature the DNA.

If conditions are brought back to normal (e.g., physiological temperature), the single stranded (denatured) DNA can renature or reform complementary strands according to the rules of base pairing (A pairs with T and G pairs with C). When the renatured DNA strands are precisely complementary, the original double stranded helical structure can reform. These same biological features of DNA make it possible to manipulate nucleic acids in vitro.

When any two nucleic acids pair together by virtue of complementarity they are said to anneal with each other and form a duplex structure. When the nucleic acids are from different sources, as occurs when one preparation consists of DNA and the other RNA, the annealing process is described as hybridization. The two common ways of performing these reactions in vitro are solution hybridization and filter (or solid support) hybridization.

Nucleic acid hybridization - the formation of a duplex between two complementary sequences, usually between two molecules that have complementary bases. It is possible for a single strand of nucleic acid that has inverted repeat sequences to hybridize back onto itself forming a stem and loop structure.

DNA - DNA hybridization
DNA - RNA hybridization

Hybrid Stability

Intrinsic factors

A duplex with relatively more GC base pairs than AT base pairs
will be more stable because there are three hydrogen bonds
between G and C and only two between A and T
Thus, it would take a higher temperature to denature

A GC base pair-rich duplex

The degree of complementarity between two strands also influences stability.

Extrinsic factors (experimental conditions)

1. temperature
2. salt concentration
3. presence of denaturing agents (e.g., formamide)
4. presence of high molecular weight polymers (e.g., dextran sulfate)

Temperature

- Ideal = 25 C below duplex melting temperature
- High temperatures may damage nucleic acids

Salt concentration

- Hybridization rate increases between 0.1 M and 1.2 M
- Commonly use 5 to 6 x SCC for solid support hybridization
  - 1 x SCC = 0.15M NaCl & 0.015M sodium citrate at pH 7.2 to 7.4
- During washing the amount of SCC is lowered depending upon required stringency

Denaturing agents

- Every 1% formamide allows lowering temperature 0.7 C without losing specificity
- Concentrations of 50% or greater formamide favor DNA-RNA hybridization over DNA-DNA hybridization

High molecular weight polymers

- Effectively increase concentration of nucleic acids by excluding volume from the hybridization mixture

Stringency

- By manipulating temperature and salt concentration, one can distinguish between perfect duplexes and duplexes that have mismatches between bases
- Under stringent conditions only perfect or near perfect duplexes can be formed
- The melting temperature of a duplex decreases 1 C for every mismatched base pair
- Relaxed conditions that allow duplex formation with mismatched base pairs include lowering the temperature

Stability of duplexes with mismatched base pairs is favored by a higher salt concentration

Wash conditions on solid supports can be adjusted to achieve the desired amount of stringency

Under stringent conditions, wash temperature can be increased and salt concentration can be decreased (down to 0.1 x SCC)

Solid Support Hybridizations

Denatured DNA or RNA is immobilized on an inert support (filter hybridization)

Prevents self-annealing

Bound sequences available for hybridization with an added nucleic acid (the **probe**)

Support filters

Nitrocellulose filters (most commonly used)

Nylon membrances (less brittle than nitrocellulose)

Cellulose paper impregnated with diazo groups

Diazo groups covalently bind to guanine residues on the DNA or RNA to stabilize support

Types of solid support hybridizations

Dot/Slot blots

Southern (DNA) **blots**

Northern (RNA) **blots**

Dot/Slot blots

DNA or RNA is bound directly to the solid support filter and then hybridized to the probe

Good for multiple samples and quantitative measurements

Specificity for qualitative measurements may be a problem for close but not identical sequences

•••

Specific Hybridization Applications

In situ Hybridization

FISH

Molecular Arrays

Hybridization Probes and Metholologies

Radioactive versus nonradioactive

Purified insert versus vector

Labeling methods

Nick translation

T4 DNA polymerase
End-labeling with T4 polynucleotide kinase
End-labeling with terminal deoxynucleotidyl transferase
End-labeling with the Klenow fragment of E. coli DNA polymerase
Random primer
Polymerase chain reaction
Riboprobes
Removal of unincorporated label after probe preparation
Use of oligonucleotides
Denatured double-stranded DNA probes

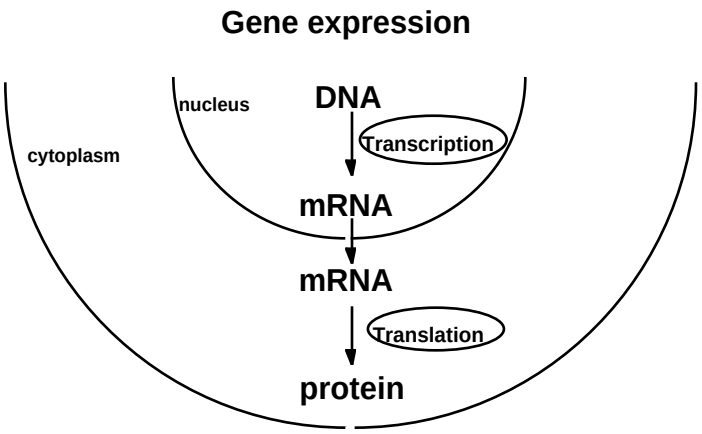
Molecular Pathology for Toxicologic Pathologists: Techniques and Applications

David E. Malarkey DVM, PhD, DACVP
Asst. Professor of Pathology
Dept. of Microbiology, Pathology, and Parasitology
College of Veterinary Medicine
North Carolina State University
Raleigh, NC 27606
work ph: (919) 513-6319 email: David_Malarkey@ncsu.edu

CONTENTS

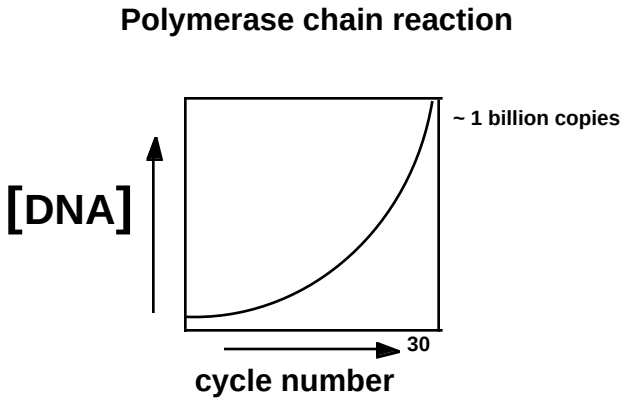
I. INTRODUCTION

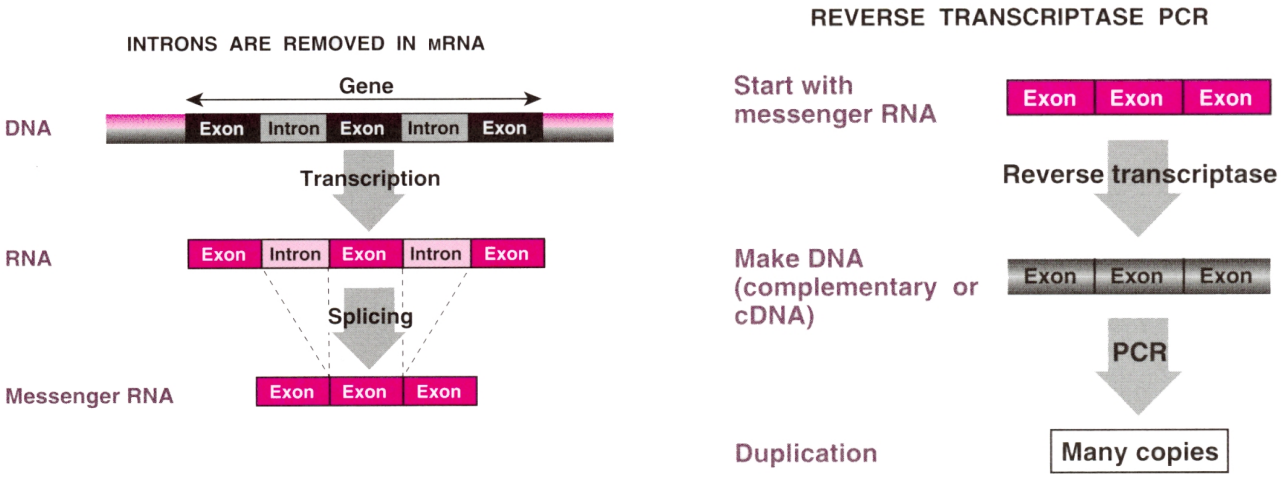
Gene expression is the process by which deoxyribonucleic acids (DNA) code, via the translation of copies of mRNA, for the production of specific proteins. mRNA copies are made from DNA in the nucleus by the process of transcription and then modified and transported to the cytoplasm where they usually become associated with ribosomes to produce protein. Gene expression can be assessed by measuring either cellular mRNA or protein. The presence of mRNA strongly suggests active gene expression, however, not all mRNA is necessarily translated into protein.



The polymerase chain reaction (PCR) and *in situ* hybridization (ISH) have revolutionized the study of genes, gene expression, and microbial infections and many of these molecular biology advances will greatly impact diagnostic pathology and toxicology research. Analysis of altered genes and gene expression can potentially provide mechanistic information about toxicity and carcinogenicity. Slide-based techniques have been developed to study mutations, gene expression, clonality, tumor cell origin, and the sequence of molecular events in cancer development in specific cell populations. This review presents a basic overview of the concepts, methodology, and applications of established and emerging molecular biology techniques in the study of DNA, RNA, and proteins.

PCR, first described about 13 years ago, is fast becoming an essential molecular biological tool in toxicologic pathology. It is a relatively simple technique to amplify specific DNA sequences to detectable and analyzable levels, and it allows the efficient generation of large amounts of DNA from minimal amounts of even crude starting material. Billions of copies of a specific molecule of DNA can be generated by reactions in less than 2 hours. PCR can amplify DNA derived from freshly frozen tissue or partially degraded tissues (such as those found in paraffin-embedded, fixed tissues) or from mRNA. For mRNA a complementary DNA (cDNA) copy is made using reverse transcriptase (RT) prior to the PCR reaction (called RT-PCR). Additional refinements of the RT-PCR technique have made it possible to make quantitative measurements of mRNA transcripts.





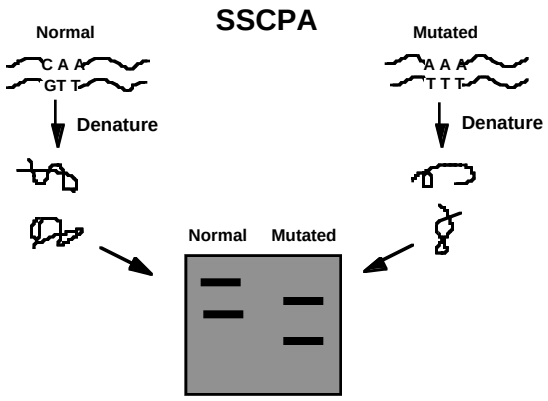
Figures from: D.P. Clark and L.D. Russell (1997) *Molecular Biology Made Fun and Simple* with permission from Cache River Press, Vienna, IIllinois, USA.

II. DNA

1. mutation analysis

PCR is commonly employed for detection of point mutations in oncogenes and tumor suppressor genes in the analysis of germline mutations and somatic mutations in tumor cells. Techniques that have commonly been used following PCR for the detection of single base changes include single-strand conformation polymorphism analysis (SSCPA), restriction endonuclease digestion analysis, allele-specific oligonucleotide hybridization (ASO), and direct sequencing.

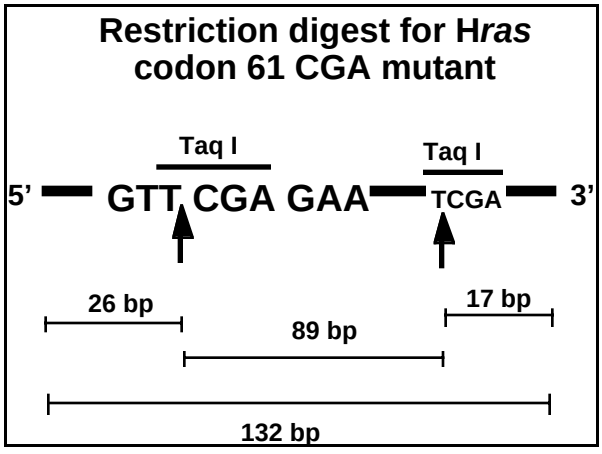
a. PCR/SSCPA



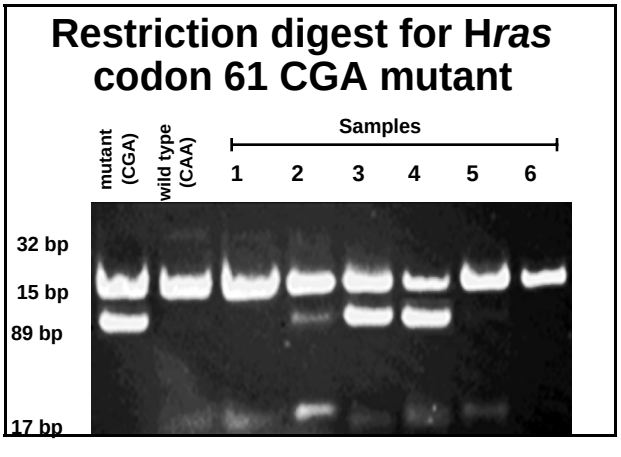
SSCPA, a quick and effective technique for the detection of single nucleotide base substitutions and the most widely used of the scanning technologies, is based on the principle that single-stranded DNA takes on sequence-based secondary structural conformation under non-denaturing conditions. Single-stranded molecules which vary by as little as a single base substitution may form a different three-dimensional configuration and migrate at an altered speed in a non-denaturing polyacrylamide gel. PCR is carried out by using oligonucleotide primers flanking the target sequence of a gene, and either the primers are labeled or, most often, a labeled deoxynucleoside triphosphate is incorporated in the reaction. Both radioactive and non-radioactive methods have been described. PCR products with altered band migration are identified on the polyacrylamide gel and further analyzed by DNA sequencing or restriction enyme digestion to define or confirm mutations.

b. PCR/RFLP

Mutations can be identified when they occur in restriction enzyme recognition sequences and render the mutated site either sensitive or resistant to cleavage. Restriction enzymes cut DNA in a sequence specific manner. For some techniques the restriction digest is used to remove the wild-type (normal) sequences, leaving only the mutated sequences intact for further analysis. Some techniques use specific primers that either recognize specific mutations or introduce a restriction site into the PCR product.



Lee and Drinkwater, 1995



Malarkey, unpublished data

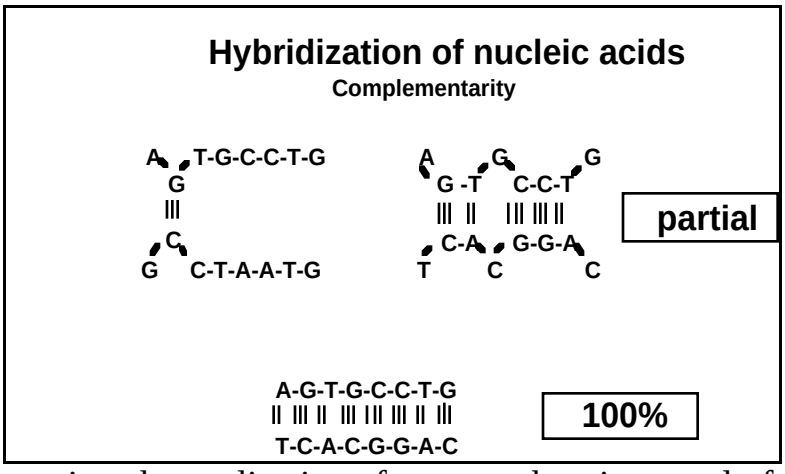
c. Allele Specific Oligoprobe hybridizations

Southern blotting is named after its originator Edward Southern and is used to identify specific restriction fragments of DNA in a mixture of DNA fragments. The complex of fragments are separated by gel electrophoresis prior to a nitrocellulose filter. The filter is incubated under hybridization with a specific radiolabled DNA probe.

Blotting Systems

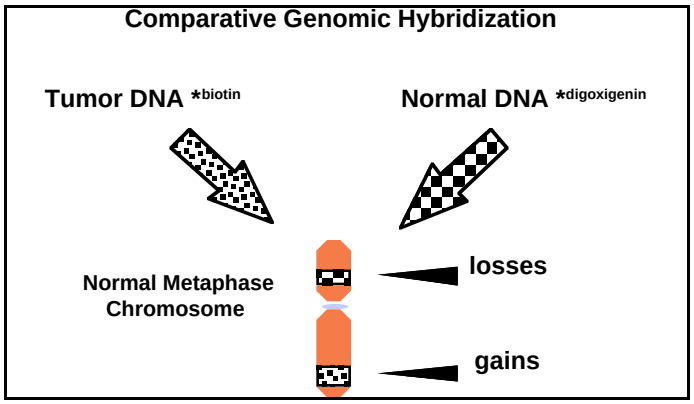
	Material on Nylon Screen	Probe
Southern	DNA	DNA
Northern	RNA	DNA
Western	Protein	Antibody
Southwestern	Protein	dsDNA

A modified method of a DNA hybridization is a technique called allele-specific oligonucleotide hybridization. Allele-specific oligonucleotide (ASO) probes have been used that can discriminate between amplification products that differ by a single base pair. The amplified target DNA is first immobilized on a nylon membrane, and then labeled sequence-specific oligonucleotide probes, usually radioactively labeled RNA probes, are applied under high stringency conditions to identify DNA products with mutant or normal sequences. Hybridization is followed by digestion with RNase which digests any DNA-RNA complexes that contain mismatches. This method often requires the application of a comprehensive panel of oligonucleotide probes to check for each possible mutation as well as the normal sequence.



d. Comparative genomic hybridization

Comparative genomic hybridization (CGH), recently developed for human cytogenetic analysis, is capable of detecting and mapping relative DNA sequence copy numbers between the genomes of two cell populations, usually between neoplastic and normal cells. CGH is based on the competitive ISH of differentially labeled tumor DNA and normal DNA



within a normal metaphase chromosomal spread. Regions of gain of DNA sequences are seen as an increased, and losses as decreased, color ratios of the two fluorochromes used to detect the labeled DNAs. Evidence indicates that many more genes than the currently known oncogenes may undergo amplification in human tumors. This technique offers promise for finding such genetic alterations in animals

e. FISH

Flourescence in situ hybridization (FISH) is useful for detecting aneuploidies, quantifying large scale chromosomal damage following exposure to clastogens, and mapping genes to specific chromosomes. It is a sensitive non-radioactive technique commonly used in human cytogenetics with great potential for application in rodents. However, there are obstacles for the use of this technique in the mouse. Success has been limited because the 40 acrocentric mouse chromosomes are of similar size and difficult to separate by flow cytometry; nonetheless, probes have been developed for the mouse.

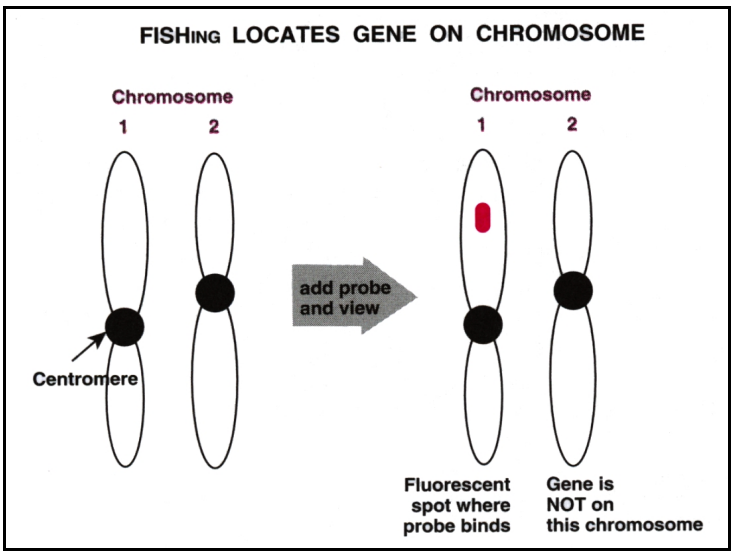


Figure from: D.P. Clark and L.D. Russell (1997) *Molecular Biology Made Fun and Simple* with permission from Cache River Press, Vienna, IIllinois, USA.

f. loss of heterozygosity (LOH)

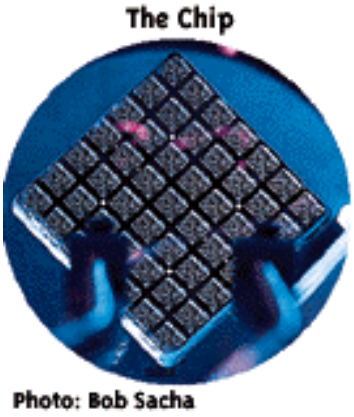
Individuals inherit one chromosomal allele from each parent, and LOH refers to the loss, by deletion or a recombinational event, of one of those alleles. Due to the recessive nature of tumor suppressor genes, mutations in these genes are frequently accompanied by an LOH in the chromosome region containing the normal allele. Allelotype analysis that reveals frequent losses of specific chromosal regions in neoplasms provides presumptive evidence for the existence of tumor suppressor gene involvement in tumor development. It is assumed that the lost or deleted chromosomal regions harbor a gene with tumor suppressor-like activity. The demonstration of LOH has been instrumental in the discovery and characterization of novel tumor suppressor genes. So far, more than a dozen tumor suppressor genes have been identified in human tumors, and many more are believed involved in both human and animal tumorigenesis.

Other than p53, little is known about the involvement of tumor suppressor genes in experimental carcinogenesis. Allelotypic analysis for LOH has recently been applied in the study of chemically induced mouse lung, skin, and liver tumors. Studies utilize F1 hybrid mice and involve PCR amplification for the analysis of

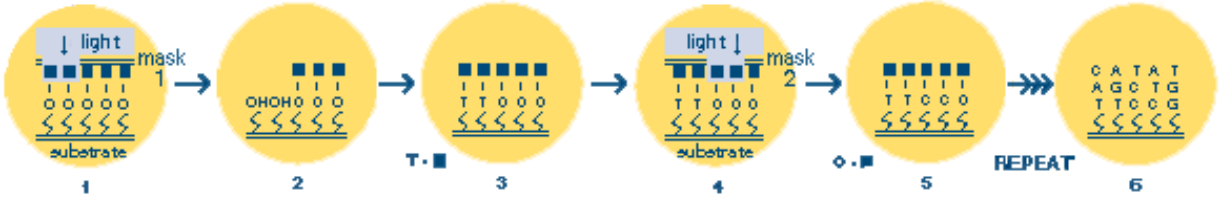
previously defined chromosomal markers, such as microsatellites. Frequent allelic losses were demonstrated in both the lung and skin tumors providing probable evidence for the involvement of tumor suppressor genes in the tumor development. Consistent LOH appears uncommon in murine hepatocarcinogenesis. As more studies are performed, the significance of these chromosomal losses and the role of tumor suppressor genes in experimental carcinogenesis may become evident.

g. DNA chip technology (see also microarray of gene expression below)

This technology has been developed to aid in the screening of polymorphisms and mutations, as well as mapping of genomic DNA clones. DNA chips will revolutionize genetic screening much the way silicon chips have transformed electronics. The computer chip makers and DNA chip makers both use a technique called "photolithography" to make microscopic "etches", however with modifications scientists at Affymetrix Inc. in Santa Clara, CA have used the technology to synthesize short chains of DNA on chip surfaces. A gene chip, which is made of glass, can hold up to 409,000 distinct testing spots each with a unique known sequence. In other words, thousands of Southern hybridizations (DNA-DNA) of an individuals DNA sample can be performed simultaneously on a glass slide about 1/2 the size of a postage stamp. Since hybridizations can theoretically be specific to find point mutations, a patient's DNA can be screened for possible mutations in cancer genes or other genes related to disease. Soon it should be possible to put a few of your cells into a gene-chip scanner and quickly analyze your genetic risks for scores of diseases. Also, tumor prognosis or specific therapeutic approaches for cancer may be dictated by the types of mutations occurring in, for example, the *K-ras*, *p53*, or *BRCA1* genes.



A gene chip begins as a glass slide about 1/2 the size of a postage stamp that contains hundreds of thousands distinct microscopic squares. Upon each square specific DNA molecules, hundreds of sequences long, can amazingly be synthesized. The highly specialized technique for DNA synthesis utilizes protecting groups on nucleic acids as "masks" to block the addition of nucleotides. Light is used to "unmask" a specific square when the appropriate nucleotide base is to be added to the molecule. DNA from the individual (can be from a tumor or perhaps peripheral



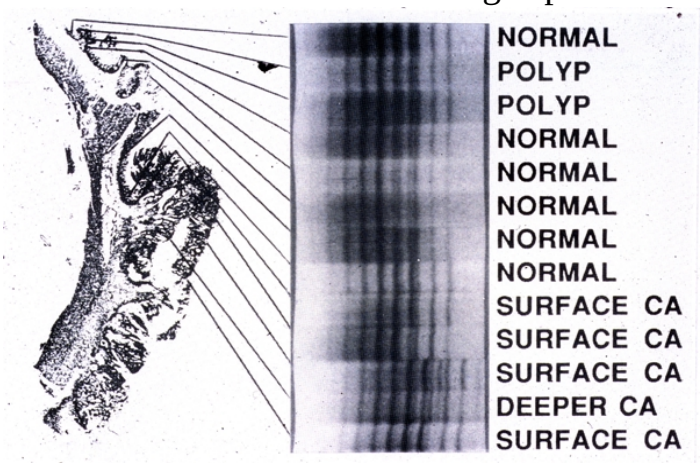
From Affymetrix, Inc. http://www.affymetrix.com/

blood leukocytes) is fragmented by restriction endonucleases and then labelled with fluorescent probes prior to hybridization. The chip is then put into an optical scanner which identifies the hybridized probes.

This same technology can be used to study gene expression (mRNA) by making cDNA copies of mRNA as described below under **microarray-based analysis of gene expression**.

h. microsatellite instability

The mammalian genome contains numerous tracts in non-coding regions of short di, tri, and tetranucleotide repetitive sequences referred to as microsatellites. For example, there are more than 100,000 CA/GT repeats that have a chain length greater than 24 bases. Investigations have reported variations in the number of repetitive unit sequences and changes in the fingerprint patterns for microsatellites in human tumors of the colon, stomach, pancreas, and endometrium. These changes were not found in adjacent normal tissue from the same individual. This and other evidence indicates that, for some cancers, a continual accumulation of mutations occurs within individual tumor cells. The evidence has led to the emerging concept that genomes of cancer cells are unstable and that microsatellites might provide a sensitive indicator for genetic instability and hypermutability in tumors. This hypothesis is based on the notion that the mutation frequency within these cells is high, and microsatellite alterations reflect the numerous mutations that have occurred throughout the genome. Loeb (1994) proposes that mutations in stability genes, for example those involved in DNA replication, DNA repair, or chromosomal segration, may be an early event in carcinogenesis and contribute to the increase in the overall level of mutations throughout the genome.



Depiction of microsatellite instability in human colorectal tumor (Shibata, et al. 1994, Nature Genetics 6:273-281.)

Alterations in microsatellites can be used as a measure of genomic instability, and recent demonstration of microsatellite instability in different human and animal tumors has provided good evidence for the so-called "mutator phenotype". Using SSCPA, PCR, and/or direct sequencing, mutations in microsatellites have been found in rat colon tumors induced by heterocyclic amines (Canzian, et al. 1994) but not in mouse liver tumors (Fox, et al. 1997). Some of the rat colonic tumors were also found to have mutations in the *APC* gene, a gene believed involved with the development of colorectal polyps and carcinomas in humans and described as familial adenomatous polyposis.

2. Diagnostics

a. Helicobacter infection (an example case).

Infection of the mouse liver by *Helicobacter hepaticus* has recently become important for the interpretation of carcinogenicity bioassays and other research utilizing rodents. Many species of the genus *Helicobacter* have been identified in mammals, and their pathogenicity varies, with some species inducing significant

disease while others appear to merely colonize the gastrointestinal tract. *Helicobacter* species identified in mice include *H. muridarum*, *H. rappini*, *H. hepaticus, and H. bilus.* Although colonizers of the mouse gastrointestinal tract, *H. muridarum* and *H. rappini* are not generally considered to be pathogenic. In susceptible strains of mice, *H. hepaticus* causes acute focal, nonsuppurative, necrotizing, hepatitis which progresses to chronic, active hepatitis characterized by minimal necrosis, hepatocytomegaly, oval cell hyperplasia, cholangitis along with the presence of relatively few infectious organisms not usually apparent on routine histologic examination. A/JCr, B6C3F1, BALB/cAnCr, C3H/HeNCr, SJL/NCr, and SCID/NCr strains of mice are variably sensitive to the development of hepatitis due to *H. hepaticus* infection, and the A/JCr and male B6C3F1 mice have an increased incidence of hepatocellular neoplasms associated with infection. *H. bilus* can be found in the bile, liver, and intestines of mice and may be associated with chronic hepatitis, although its pathogenicity is, as yet, undetermined, and it is not known whether *H. bilus* causes liver tumors in mice.

Murine Helicobacters

Agent	species	disease
H. hepaticus	mouse	hepatitis, liver tumors, IBD*
H. bilus	mouse / rat	hepatitis / IBD*
H. muridarum	mouse	gastritis
H. trogontum	rat	-
H. rodentium	mouse	-
"F. rappini"	mouse	-

*IBD=inflammatory bowel disease / typhlocolitis

There are multiple approaches to identifying a study (or group of mice) as being infected. Definitive diagnosis of *H. hepaticus*-associated liver disease in mice is based on the presence of histologic alterations along with demonstration of the infectious agent. While culturing organisms from fresh or frozen tissue or feces is desirable, retrospective analysis will depend upon alternative methods, including PCR-based assays, that can be applied to archival tissues. *H. hepaticus* organisms have been detected in lesioned liver by methods such as Steiner or Warthin-Starry histochemical stains, immunoflourescence, *in situ* hybridization, immunohistochemistry, and ultrastructural analysis. A serum ELISA test for circulating IgG antibodies against *H. hepaticus* has been used to provide information for determining *in vivo* infection status in pathogenesis studies. PCR-based diagnostic assays for *Helicobacter hepaticus* potentially offer greater specificity, increased sensitivity, and decreased expense and time expenditure compared to culture, histochemical stains, or ultrastructural analysis. Recently a number of PCR-based assays have been described for the identification of murine helicobacters in rodent tissues or feces; and

PCR detection of *H. hepaticus*

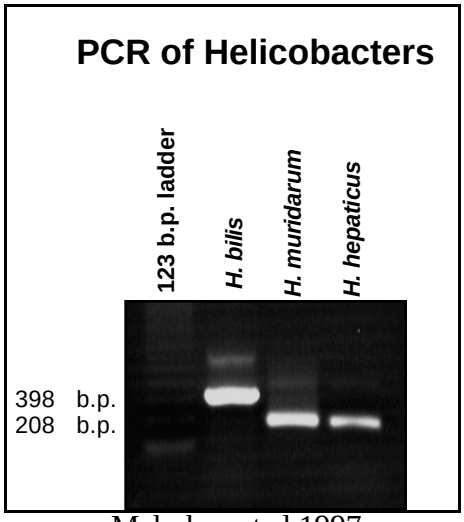
Chemical	Helicobacter hepatitis	Steiner stain	PCR/RFLP
Triethanolamine (5)	80%	80%	100%
AZT (20)	35%	10%	55%
Scopolamine (20)	5%	5%	60%

Malarkey, et al. 1997. Tox Path 25 (6):606-612

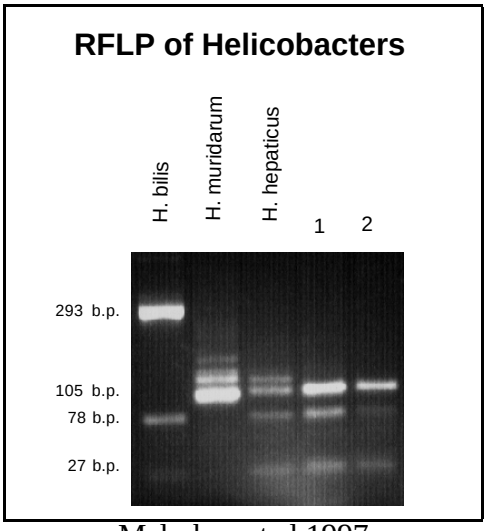
some have reported success at detecting *H. hepaticus* in formalin-fixed, paraffin-embedded tissues.

Presently, the specificity of many PCR-based assays for the identification of the Campylobacteria family, which includes *Helicobacter sp.*, should be regarded with caution. Specificity of PCR-based molecular tests are dependent on the degree of variability in the primary nucleotide sequences of the 16S ribosomal RNA (rRNA) gene. These sequences are nearly identical in phylogenetically closely-related bacteria. For example, the region of the 16S rRNA gene amplified in the assay described by Malarkey, et al. (1997) for *H. hepaticus* has similarities with other Helicobacter species not reported in the mouse, such as *H.canis, H. trongontum, H. mustelae,* as well *H. rappini*, found rarely in the mouse and not considered a murine hepatic pathogen, making them indistinguishable by the PCR-RFLP. It is possible that as newly discovered species of *Helicobacter* are characterized, some may have spans of nucleotide sequences indistinguishable by PCR from that of *H. hepaticus*. Confidence in the diagnosis of *H. hepaticus* hepatitis can be gained by using multiple PCR-based assays at different regions of the 16S rRNA gene to distinguish *Helicobacter* organisms based on sequence polymorphisms.

Diagnostic PCR amplification of genes for the identification of presumptive low level bacterial infection from formalin-fixed paraffin embedded archival tissues has proven problematic presumably because fixation is inherently detrimental to nucleic acid molecules and the PCR. The determinants of sensitivity of the PCR assay appear related to the bacterial load and distribution in the liver, amount of tissue analyzed, and the detrimental effects of formalin fixation and paraffin embedding procedures on the bacteria in the tissue and the PCR reaction.



Malarkey et al 1997



Malarkey et al 1997

Shortening the duration of fixation, increasing the initial sample size, amplifying relatively small fragments (less than 200 base pairs), using a nested primer amplification method, and/or increasing the number of amplifying cycles have all been shown to be helpful strategies for achieving successful PCR results from fixed paraffin-embedded tissues. In the studies *H. hepaticus* infection in mice by Malarkey, et al. (1997), results in liver tissues appeared primarily dependent on the duration of fixation and bacterial load, and not necessarily on the presence or absence of hepatitis, origin of the sample from neoplastic or non-neoplastic liver, or sex of the mouse. In some cases PCR methods were more sensitive than histologic evaluation and thus may provide the first evidence of *H. hepaticus* infection. The finding of PCR positivity in the absence of histologic alteration may be due to identification of an early infection or of a strain of *H. hepaticus* that is of low or minimal pathogenicity.

PCR detection of *H. hepaticus*

		PCR-RFLP results	
Chemical	Helicobacter hepatitis	Frozen liver	fixed liver
Triethanolamine (5)	80%	100%	80%
Cobalt Sulfate (20)	55%	NA	0%
AZT (20)	35%	55%	10%
Scopolamine (20)	5%	45%	40%

Malarkey, et al. 1997. Tox Path 25 (6):606-612

In summary, the impact of infection of the mouse liver by *Helicobacter hepaticus* on research and the bioassay is not completely understood and our understanding of the pathogenesis depends on achieving a reliable diagnosis.

III. mRNA

a. northern blotting

Northern blotting is the RNA counterpart to Southern blotting and is used to detect a particular RNA in a mixture of RNAs. The differences between the two procedures stem from the fact that RNA is single stranded and must be denatured, ensuring that all RNA molecules have an unfolded, linear conformation, for effective electrophoresis. The individual RNAs are separated according to size by gel electrophoresis and transferred to a nitrocellulose filter. The filter then is exposed to a labeled DNA probe and subjected to autoradiography. The assay will assess the steady state level of specific mRNAs at the time the RNA was extracted from cells or tissue.

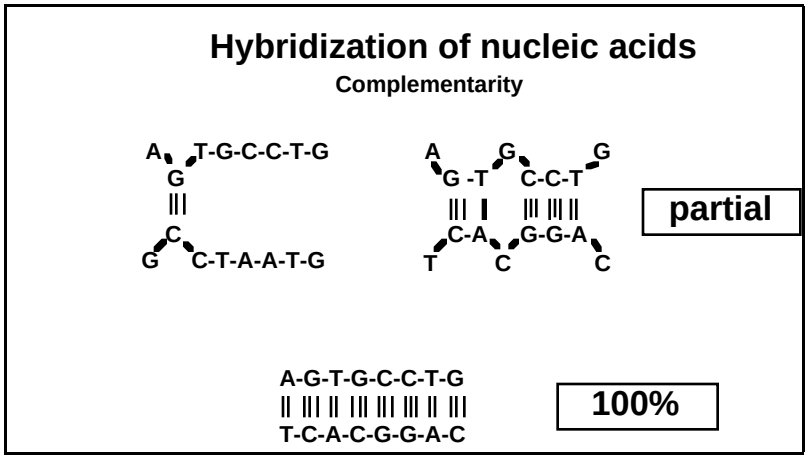
b. Rnase (nuclease) protection assay

A method for detecting and quantitating specific RNA molecules based on the principle that endonucleases digest single-stranded but not double-stranded nucleic acids. In the technique, a radiolabeled ssDNA probe (added in excess) specifically hybridizes with complementary target RNA in a mixed population of cellular RNAs and is then subjected to digestion with an endonuclease. The endonuclease degrades all the unprotected ssRNA and DNA molecules leaving the double-stranded RNA-DNA complex intact. The protected RNA-DNA hybrid is detected by

autoradiography after gel electrophoresis and the amount of intensity of the resulting band is proportional to the original quantity of RNA in the sample.

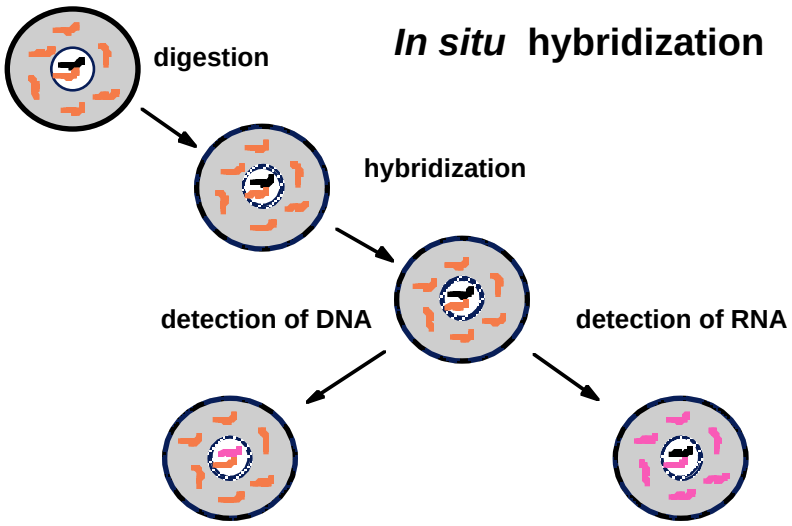
c. ISH

ISH, described over 25 years ago and only recently widely utilized, can localize specific sequences of DNA and RNA within cells and thus provide data on individual cells rather than an average of total cellular populations. ISH is based on the principle that cell and tissue-bound RNA and DNA sequences will hybridize with labeled probes of complementary sequence. Sites of hybridization can then be visualized microscopically. ISH also allows the maximal use of tissues, through serial sections, that may be of limited supply. The technique can be extremely sensitive and detect the relative amounts of mRNA contained in a single cell. Over the last decade there have been numerous refinements of ISH technology that have allowed for some new applications in basic cancer research.



ISH has been used for the detection and localization of gene expression, microbial infections, single-copy genes to individual chromosomes, translocations, deletions, amplifications, and other chromosomal abnormalities. ISH has proven a useful tool in the study of c-H-*ras* and c-*myc* expression in chemically induced liver tumors in B6C3F1 mice. In conjunction with immunohistochemistry and PCR analysis, ISH can add to our understanding of the role of oncogenes in tumorigenesis, as described in studies of *myc* in human malignancies.

Methodology. The success of this technique depends on many factors including the number of copies of DNA or RNA present, extent of preservation or fixation, type and size of probe, stringency conditions for hybridization, ionic strength of reaction, proportion of the nucleotide bases G and C, proportion of mismatch bases, method of probe labeling, and techniques of signal



detection. Some protocols are claimed to have a 90% success rate on the first hybridization attempt despite the source or type of tissue.

Of importance is the inclusion of proper controls including cells known to express the target gene, non-expressing cells, use of prehybridized probes, hybridization with non-specific vector sequences, and use of corresponding sense probes that should not hybridize to the specific mRNA.

Fixation. The evaluation of many fixatives for ISH has revealed that 4% paraformaldehyde seems to achieve the best overall results. Excellent results have been attained, however, with many of the cross-linking aldehyde fixatives, including formalin-fixed, paraffin-embedded tissues. Immediate quick fixation is recommended to inactivate RNA degradation by endogenous ribonucleases. Fixatives such as Bouin's, Zenker's, and Carnoy's should be avoided because they generally provide less preservation and retention of mRNA. When mRNA is abundant the length and time of fixation is less critical. Fresh frozen sections have also been used successfully with quick fixation from a few minutes to a few hours. The use of frozen sections for ISH, despite imparting a loss in morphologic detail, obviates the need for deparaffinization and protease digestion -- steps that might contribute to false negative results.

Pretreatment of slides. Pretreatment of slides with aminopropyltriethoxysilane, poly-l-lysine, or organosilane increases adherence of tissue sections. Following deparaffinization, protease digestion (proteolysis) is performed with using enzymes such as proteinase K, pepsin, or pronase to unmask the target nucleic acids and allow penetration of the probe. The degree of protease digestion is related to the length of fixation and tissue type; generally the longer the fixation, the greater the amount of digestion required. For some protocols treatment with acid solutions and/or excessive heat are required to achieve optimal results. A balance between adequate digestion and retention of morphology must be determined for each tissue type, probe, and reaction. DNase or RNase digestion may be performed to remove nontarget nucleic acid sequences.

Probe selection. Probes can be single-stranded (ss) or double-stranded (ds) DNA or RNA and ideally should be less than 400 bases long to permeate the cell matrix. ssRNAs are preferred for most experiments because RNA:RNA hybrids are more stable, and digestion with RNase removes unhybridized and partially hybridized ssRNA, thus decreasing non-specific, false positivity. dsDNA probes are most commonly used for the detection of DNA. Oligonucleotide probes (oligoprobes) of 20-40 bases of ssDNA are usually made by chemical synthesis, while longer ssDNA probes are made from clones or unidirectional PCR. The production and use of oligonucleotide probes does not require advanced expertise with basic molecular biology technologies because cloning technology is not necessary, labeling can be carried out in a single step, and oligoprobes can be constructed based on published cDNA maps and are increasingly available from commercial sources. Oligoprobes are somewhat less sensitive than ss or dsDNA probes with longer sequences because relatively few labeled nucleotides can be incorporated per molecule of probe; however, sensitivity is increased by using cocktails of up to 20 different specific oligoprobes simultaneously. The attainment of high specificity under stringent conditions is possible with oligoprobes.

There are multiple ways to label the probes using either radioactive (^{35}S, ^{32}P, or ^{3}H) or non-radioactive (biotin, alkaline phosphatase, digoxigenin, flourescein) labels. In general the radioisotopic labels are best for quantitative analysis and have increased sensitivity and decreased resolution compared to non-isotopic methods. However, in the last ten years dramatic advances in non-isotopic labeling and detection systems have greatly enhanced their sensitivity. These systems are yielding a high resolution and decreased non-specific background allowing better localization of the target. Non-isotopic methods are relatively simple and nonhazardous, have a shortened procedure and turn-around time, and occasionally have sensitivity equivalent to that of radioisotopic methods.

Applications of ISH

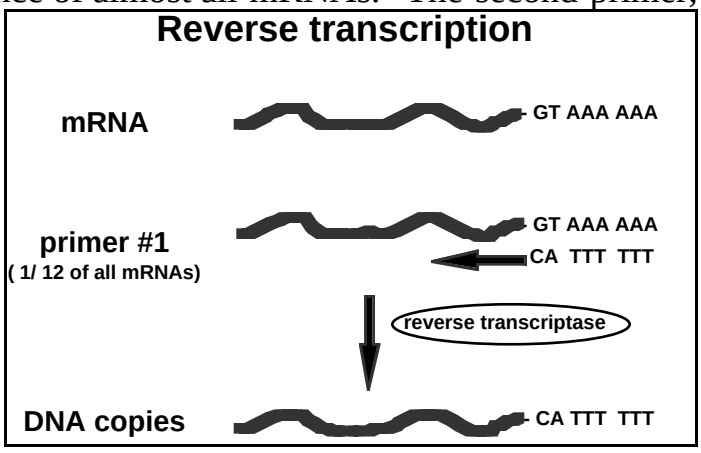
Oncogene/tumor suppressor gene overexpression. ISH has proven a useful tool in the study of *Brca1*, c-H-*ras* and c-*myc* expression in chemically induced liver tumors in B6C3F1 mice (Malarkey et al., unpublished results; Nelson, et al. 1990). In conjunction with immunohistochemistry and PCR analysis, ISH can add to our understanding of the role of oncogenes in tumorigenesis, as described in studies of *myc* in human malignancies (Tervahauta, et al. 1992).

Measurement of cell proliferation. The expression of histone genes is known to be tightly coupled to DNA synthesis, and the presence of histone mRNA appears to be a reliable marker of cells in the S-phase fraction. Detection of histone mRNA is proposed to be more accurate than proliferating cell nuclear antigen (PCNA) immunohistochemistry as an endongenous marker for S-phase cells (Alison, et al. 1994), and examination of archival tissue is possible. Because the histone genes are conserved between species, the same probes can be applied to human and animal tissue. Determination of histone mRNA has been used for measuring cell proliferation in normal and malignant human tissues (Chou, et al., 1990) and in regenerating rat liver after partial hepatectomy (Alison, et al. 1994).

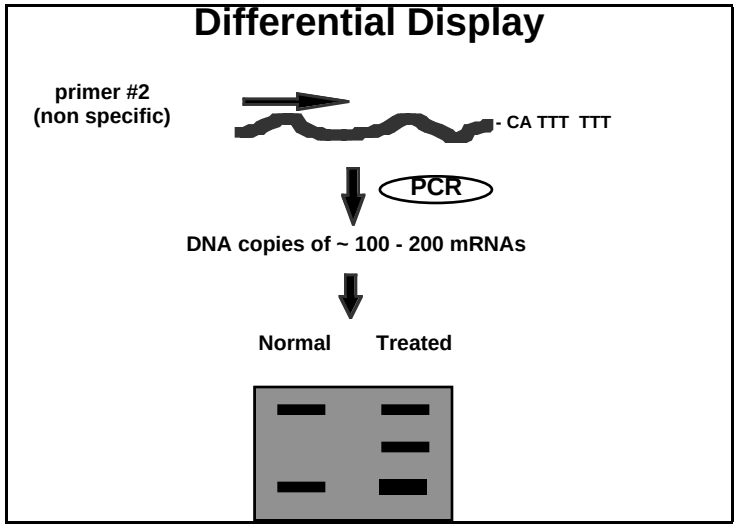
d. differential display of mRNAs

Comparisons of gene expression in different cell types can provide underlying information for understanding biologic processes that dictate development and disease. The analysis of changes in gene expression (either induction, enhancement, or suppression) in cells which have been exposed to certain chemicals or undergone neoplastic transformation is of prime interest in toxicological pathology. The technique called differential display of mRNA allows identification of new genes or the characterization of differentially expressed genes by comparing expression between two samples. Differential display offers a marked savings of time over cloning-based or subtractive hybridization techniques in searching for new genes, allows simultaneous detection of differentially expressed genes between different cell populations, is a simplistic method initially requiring only PCR apparatus and a sequencing gel, and requires only a small amount of RNA.

The aim of differential display is to amplify partial cDNA sequences from subsets of the approximately 15,000 mRNAs in any given cell using RT-PCR. Primer pairs are designed to amplify cDNA copies of 50-100 mRNAs at a time. The first primer anneals at the junction of the polyadenine tail which is naturally added on to the 3' untranslated sequence of almost all mRNAs. The second primer, an arbitrary primer of about 10 nucleotides of defined sequence, is added to the reverse transcriptase mixture along with polymerase and a radioactive nucleotide prior to PCR amplification. Amplified fragments are separated on a denaturing gel, and fragments present in one sample and absent from the next are determined by inspection of a suitably exposed autoradiograph. Once identified, DNA is cut directly from the gel and available for reamplification, Northern blotting, cloning, sequencing, and further characterization.



Novel mammalian genes have been identified and characterized utilizing differential display. The technique has been used successfully to find a novel gene induced by dioxin in rat liver. Further characterization of that gene has led to the discovery that it has homology with an, as yet, undefined human gene. Furthermore, investigators demonstrated a dose-related increase gene expression in rat liver and constitutive expression in other organs including brain, lung, and kidney . Others researchers using the differential display technique have identified genes expressed specifically in human brain tumors or rat brain after treatment with cocaine or amphetamine. As the technique is refined further and more widely used, information will be gained regarding genes involved with chemical toxicity and carcinogenicity.



Differential display of mRNAs

•Aftermath

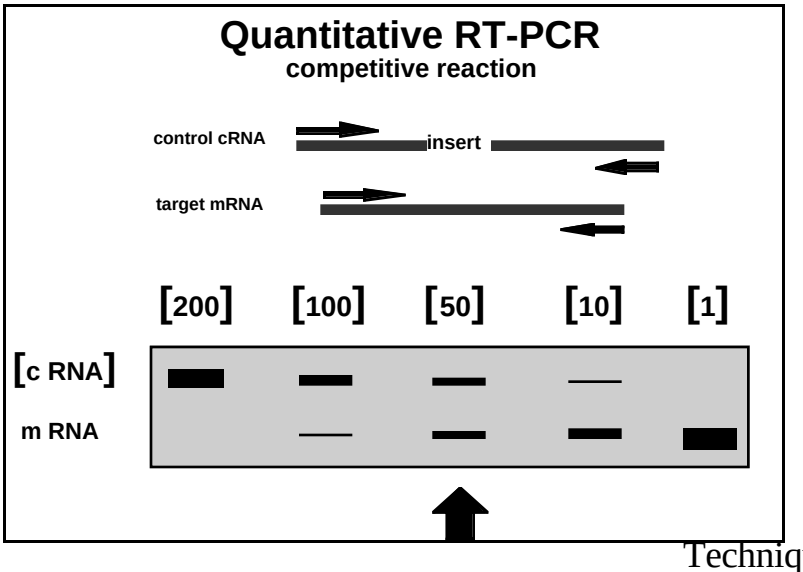
- Characterize the gene
  - DNA sequencing
  - check cross-species homology
  - analyze expression in other organs and effects of treatment

e. quantitative reverse transcriptase PCR (QRT-PCR)

If RNA instead of DNA is used as the initial sample, producing a DNA copy of the RNA using reverse transcriptase before starting the PCR becomes necessary. Reverse transcriptase will generate cDNA from mRNA. Some of the commercially available polymerases function both as a reverse transcriptase and a DNA polymerase allowing the production of the first strand cDNA and cDNA amplification in the same tube. Assessment of gene expression by PCR is proving to be an extremely useful technique, especially in situations in which only a limited number of cells is available. Additional refinements of the RT-PCR technique have made it possible to make quantitative measurements of mRNA transcripts. QRT-PCR is full of both great potentials as well as problems. The results are dependent on the efficiency of the amplification and variability can be very large, since PCR can also easily amplify errors, and thus preclude accuracy and reliable quantification. Furthermore, none of these approaches control for the variability present in the RT reaction. The techniques must be carefully designed and executed for each experiment and lab in order to achieve reliable results.

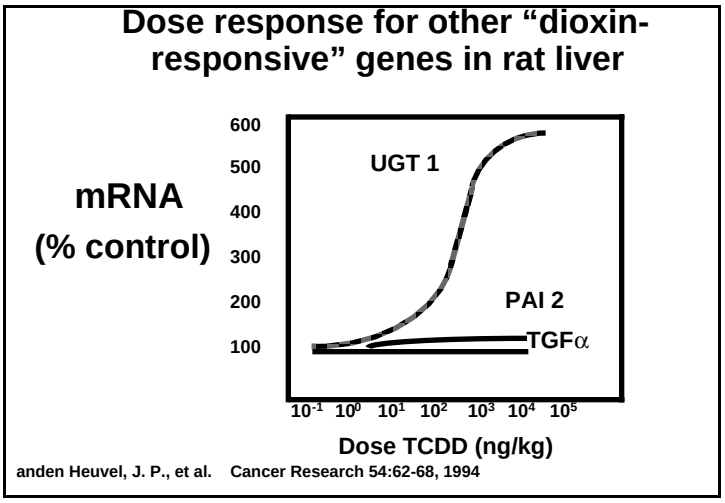
i.competitive and non-competitive QRT-PCR

Although PCR has had a profound impact in many areas of research, its application to the quantitation of nucleic acids has proven problematic. This difficulty is due primarily to the exponential nature of PCR, as small variations in the amplification efficiency dramatically affect the yield of amplification product. Despite this limitation, absolute quantitation of RNA can be readily achieved using a technique called competitive RT-PCR, which involves the spiking of individual reactions with known amounts of control RNA template that is coamplified with the target of interest. The control RNA (cRNA) is made by in vitro transcription and is identical to the target RNA except for a small deletion, insertion, or restriction site introduced by mutagenesis. The control



template is amplified by the same primers as the endogenous target to minimize differences in amplification efficiency. After a standard concentration curve is generated by amplifying serial dilutions of spiked control RNA, direct quantitative comparisons between RNA levels can be made. The point at which the concentration of target and control are equimolar (arrowhead) can be used to calculate the original quantity of target RNA. Endogenously expressed RNA can serve as an internal control.

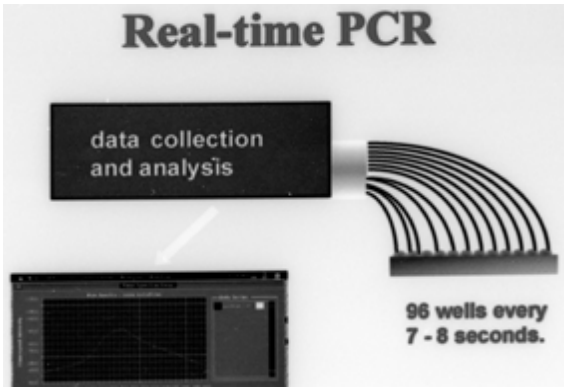
This methodology has been used to study dose-response relationships of dioxin-responsive genes in the rat liver (Vanden Heuvel, et al. 1994). The investigators found dose-related increases in the expression of the enzymes cytochrome p450-1A1 and UDP-glucuronosyl-transferase-1 (UGT 1) but no increases in transforming growth factor-a (TGFa) or plasminogen activator inhibitor-2 (PAI2) at four days following a single exposure to dioxin. This type of data adds to our understanding of the pertinence of the dynamic response of gene expression for certain environmental chemicals that pose potential human health hazards. The amount of gene expression may also be used as a biomarker of human exposures to environmental chemicals (Vanden Heuvel,et al. 1993). Some advantages of this technique include the benefit of examining many genes concurrently and the ability to quantify absolute levels of a given mRNA and attain a high sensitivity of detection.



Noncompetitive RT-PCR differs only in that the quantification is estimated based on conditions in which there is no competition for the components of the PCR. The standard RNA is added in a linear fashion within one order of magnitude of the target. The increasing series of standard amounts is co-amplified with equal amounts of total experimental RNA which generates a linear scaled graph. The intercept point (between native and standard) is then used to estimate quantity of the target mRNA.

ii. real-time (kinetic) QRT-PCR

Real-time QRT-PCR is a relatively new approach that is based on the measurement of the target DNA produced during each cycle of an amplification reaction. The traditional amplification product measurements have been the "end point" analysis which are the determinations made after it is completed. Real-time product monitoring offers the greatest potential for improved QRT-PCR.

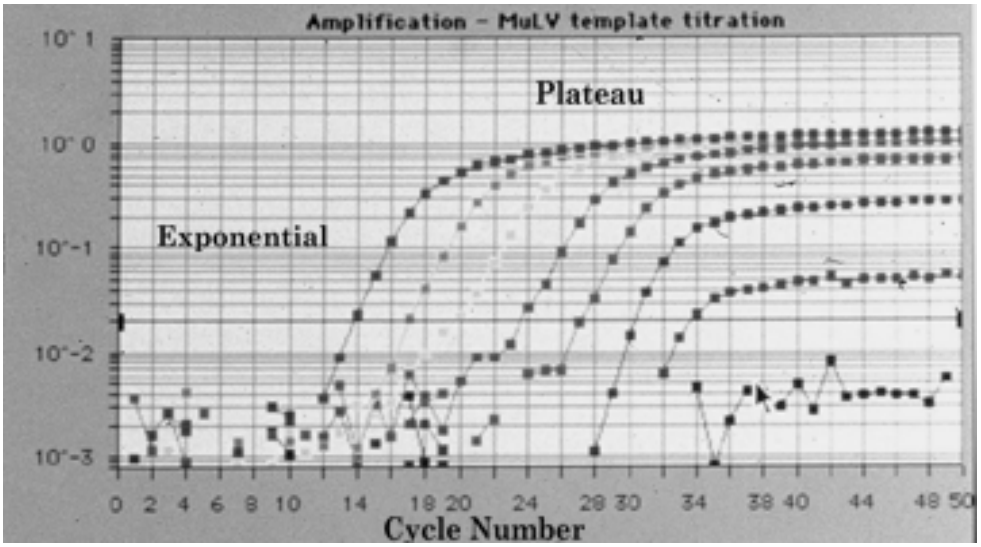


Two methods have been designed to provide real-time detection of amplified products, the ABI PRISM 7700 and 5700 Sequence Detection Systems (Perkin Elmer Applied Biosystems) and the LightCycler™ (Idaho Technologies). The quantitative ability of this detection method comes from being able to monitor the accumulation of amplification products using either fluorogenic probes or intercalating dyes. The methods are based on the kinetics of the PCR reaction and the fact that, under optimal conditions, the cycle number is proportional to the amount of starting material. The PCR occurs in two phases, the exponential phase and the plateau phase. The exponential phase occurs during the early and middle cycles and the amount of cycles before it enters this phase is dependent on the amount of starting cDNA. Relative differences in the number of cycles required to reach the midpoint of the exponential phase can be used to mathematically calculate (based on controls of known amounts of cDNA) the starting concentrations of cDNA. During the plateau phase the components of the reaction mixture (ie- supply of nucleotides, polymerase activity, primer concentration) become limiting. Furthermore the numerous single stranded products may re-anneal with each other rather than with the primers.

The PRISM 7700 system uses probes (TaqMan; Perkin Elmer) to specifically detect the target sequence in the presence of nonspecific amplification products. A specific flourogenic probe, that hybridizes internally to the amplified product, is added to the reaction. The 5'-> 3' exonuclease activity of the *Taq* polymerase hydrolyzes the probe which abolishes the suppression of the reporter flouescent dye. The flourescence emission is measured during each cycle through a fiber optic lines positioned above an optically non-distorting tube caps. The amount of flourescence detected is proportional to the amount of accumulated PCR product.

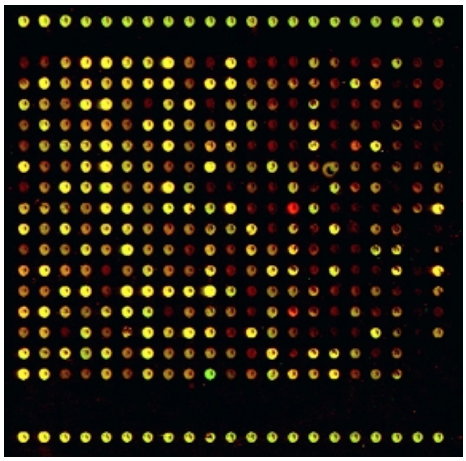
f. microarray-based analysis of gene expression

Microarray technology is one of several developing approaches to comparatively analyze genome-wide patterns of mRNA expression. Simply stated, the recently described cDNA microarray or DNA chip technology allows for the simultaneous monitoring of the relative expression levels for hundreds to thousands of genes between two populations of cells (ie between tissue from treated and control animals or between an individual's diseased and normal tissues).



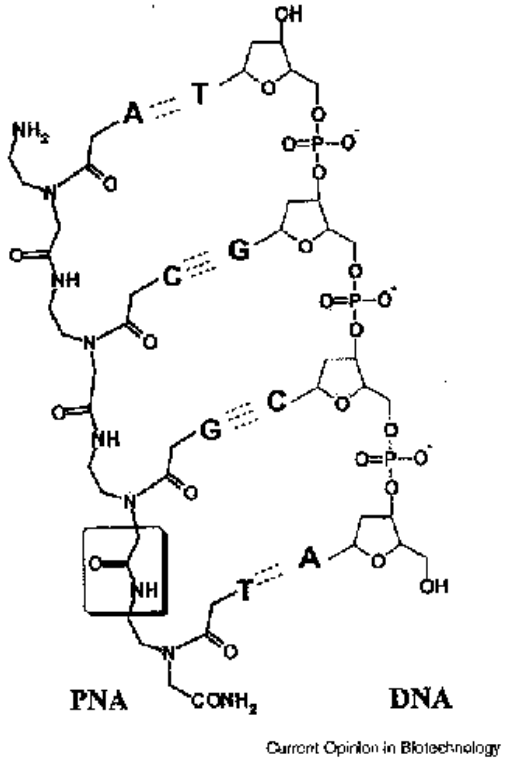
The technology uses specific complementary cDNA sequences or cDNA inserts of a library for PCR amplification that are arrayed as "microdots" on a glass slide. Each "microdot" represents a pure population of a specific cDNA and with high speed robotics as many as 1000 cDNA sequences are arrayed per cm^2. The microarrays serve as gene targets for hybridization to cDNA probes made from RNA samples of cells or tissues. A two-color fluorescence labeling technique is used to label the probes and during simultaneous hybridization fluorometric signals reflect the relative abundance of specific gene expression. The ultimate goal is to develop arrays which contain every gene in a genome against which mRNA expression levels can be quantitatively assessed.

In the laboratory of Dr. Cynthia Afshari at the NIEHS, RTP, NC (http://www.niehs.nih.gov/envgenom/abstract/z01smart.htm), microarray technology is being developed to identify toxicants and carcinogens in humans and rodents. The patterns of gene expression may reveal data that lead to assessing exposures as well as elucidating the mechanisms of action for environmental agents. For example, treatment of model systems (such as dioxin, peroxisome proliferators, or estrogen) may provide a "molecular signature" that can be used to gain insight into affects of unknown agents or determine crosstalk between combinations of agents.



IV. PNA (Peptide Nucleic Acids)

PNA, as first described by Nielsen, et al in 1991, was originally designed as a reagent to specifically bind double stranded DNA in order to control gene expression. The intent was to develop targeted antisense and antigene therapeutic drugs. PNA is an analog of DNA in which the entire negatively-charged sugar phosphate backbone is replaced with a peptide-like backbone. The peptide backbone consists of repeated units of N-(2-aminoethyl)glycine linked by amide bonds and this backbone supports the four natural nucleotide bases of adenine, cytosine, guanine, or thymine which are at spacing equal to that of the DNA bases.



Nielson,PE. 1999.Curr Opin Biotech 10:71-75

There are a number of unique structural and hybridization properties of PNA that offer many potential biological and diagnostic applications than that of traditional oligonucleotides. For example, PNA is not prone to degradation by nucleases or proteases, thus offering high biological stability. PNA, having a neutral backbone and proper interbase spacing, binds to its complementary nucleic acid sequence (DNA or RNA) according to Watson-Crick base-pairing rules with higher specificity and affinity. Furthermore, the stability of the PNA/DNA duplex is essentially independent of salt concentration in the hybidization solution. There is also a higher thermal stability of the PNA/DNA duplexes, and this stability is strongly affected by imperfect matches. PNA/DNA duplexes with a mismatch is more destabilizing than a mismatch in DNA/DNA duplex. This property of discrimination makes PNA attractive for use in assays to detect specific mutations in DNA molecules.

PNA technology exploits the unique properties of PNA and is emerging as a breakthrough new technology which appears destined to revolutionize genetic diagnostics, therapeutics, and the study of gene function. Techniques are currently being developed to utilize PNA in PCR, arrays, mutation detection, in-situ hybridization (ISH), flourescence in-situ hybridization (FISH), selective suppression of wild-type genes in the detection of low levels of oncogenes, biosensors in conjunction with mass spectrophotometry, gene therapy, and diagnostics.

V. Analysis of proteins

Western blotting

Method for the detection of a particular protein in a complex mixture that utilitizes a 3 step procedure of gel electrophoresis, the specificy of antibodies, and the sensitivity of enzyme assays. The protein mixture is first separated on an SDS-polyacrylamide gel. Then a thin nitrocellulose membrane is applied to the face of the gel to bind and transfer the proteins. An electric field drives the proteins out of the gel and into the membrane (a.k.a.= blotting). The membrane is soaked in a solution containing the specific antibody to the protein of interest and the band is visualized after treatment and development of an enzyme-linked secondary antibody.

immunohistochemistry

The immunohistochemical detection of specific proteins on tissue sections is the pathologists' forte (as well as the bane of our existence). Don't forget to use proper controls and test the antibody's specificity!!!!!!

IV. slide-based techniques

Slide-based PCR techniques have been recently described and are being developed to allow the analysis of DNA or RNA from small subpopulations of cells in a tissues section. Heterogeneity of mutations among neoplastic cells has been observed in both human and animal tumors. Some studies describe techniques with careful microdissection of small areas of tumors from histologic sections followed by a PCR-based mutation detection assay.

a. ISH (see above)

b. Microdissection of tissue

New slide-based microdissection techniques have recently been described and applied in the analysis of DNA, RNA, and/or protein from small subpopulations of cells in tissue sections, and in some investigations have helped delineate the sequence of molecular events of human tumor development and progression. The precision of microdissection varies among the techniques, and, until laser capture microdissection became available in 1997, methods have been capable of analyzing no fewer than about 50 cells from a histologic section. The results will aid in our understanding of the sequence of mutational events and gene expression in tumor development and progression, and futhermore may aid in studying clonality, tumor diagnosis, and prognosis of neoplasms.

i manual scraping

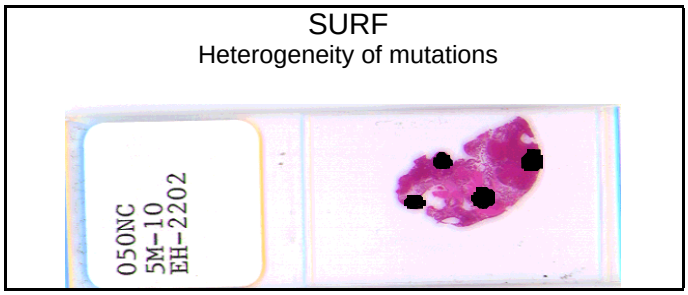
Be careful not to contaminate. Need we say more?

ii. glue

Recently a technique using a starch-based adhesive has been sugested as an alternative for the selection and study of small subpopulations of cells (Turbett, et al. 1996). Briefly, a starch adhesive glue was placed over the areas of interest as small as 1 mm^2 and the dried spots of glue were removed along with the tissue and placed into a tube for subsequent DNA isolation. Similar techniques are currently being applied by some investigators for the study of mRNAs.

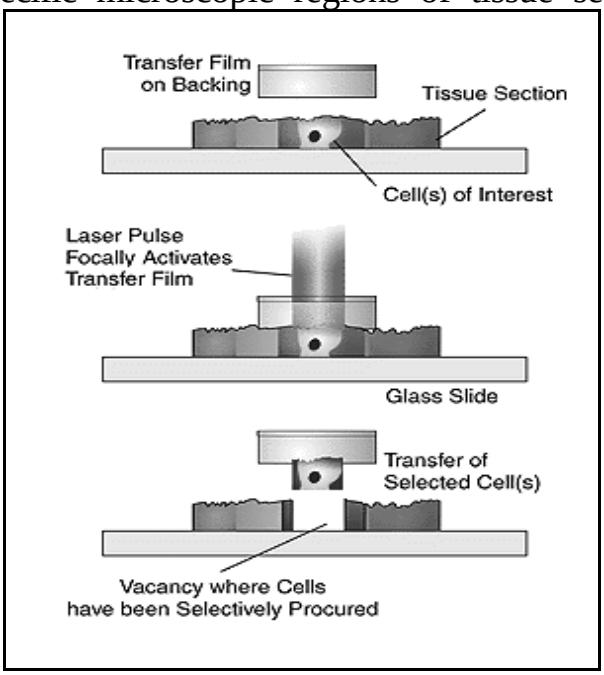
iii. SURF

A specialized technique called selective ultraviolet light radiation fractionation (SURF) allows for the selection of a subpopulation of as few as 30 cells from surrounding tissue in a histologic section (Shibata, 1992). Ink dots are applied over the tissue to protect subpopulations of tumor cells from short-wave ultraviolet light exposure. The unprotected cells are damaged so that DNA targets are not amenable to PCR amplification. After the appropriate exposure to ultraviolet light the protected cells are then cut away from surrounding tissue, which is mounted on plastic slides, and used as the DNA source for mutational analysis.



iv. Laser capture microdissection

Laser capture microdissection is a promising new method for procuring as few as 2-3 pure cells from specific microscopic regions of tissue sections with precision and without contamination (Emmert-Buck, et al 1996). The microdissection is accomplished when a laser beam focally activates a special transfer film which bonds specifically to the targeted cells identified by microscopy within the tissue section. The transfer film with the bonded cells is then lifted off the thin tissue section, leaving all unwanted cells behind. Multiple homogeneous samples can then be analyzed by PCR.



website: http://dir.nichd.nih.gov/lcm/lcm.htm

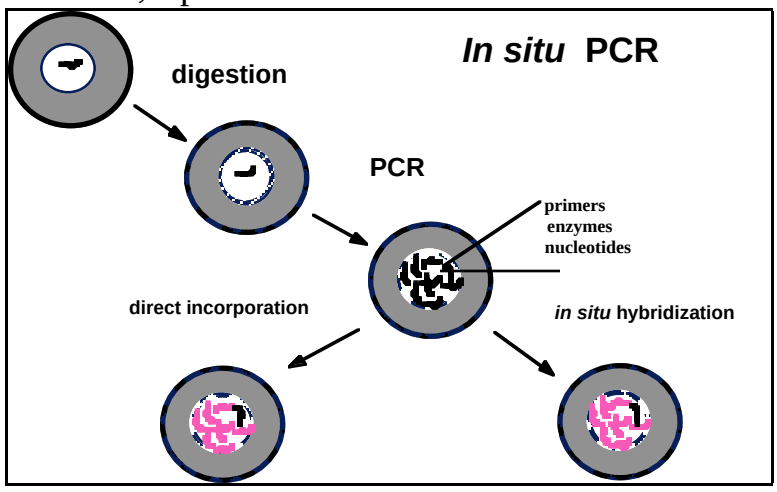
c. IN SITU PCR (IS-PCR)

By combining the localizing ability of ISH with the amplifying capability of PCR, the technique of IS-PCR has allowed the detection of low copy numbers of specific DNA and mRNA sequences in tissue sections and cell preparations. IS-PCR has been successful for the detection of low copy number viral DNAs, single-copy genes, immunoglobulin gene rearrangements in cell suspensions and cytospins, and for the analysis of low-level gene expression. The greatest success has been gained in the study of infectious diseases and, in particular, in the diagnosis of viral latency. The advantages of IS-PCR include elimination of concern over contamination-based false positives, ability to assess the histologic and subcellular distribution of the target, and ability to detect as few as 1 -10 copies in a cell. IS-PCR also has been extended to studies of RNA after the *in situ* synthesis of the corresponding cDNA through the use of reverse transcriptase. This approach has been used successfully for the detection of RNA viruses and for mRNAs.

Detection threshold of major hybridization assays

	detection threshold
In situ PCR	1- 10 copies / cell
In situ hybridization	> 10 copies / cell
Southern blot	1 copy / 100 cells
Slot blot	1 copy / 200 cells

The procedure seems straightforward; however, many scientists have been unable to achieve reproducible results. Results from work with archival formalin-fixed, paraffin-embedded materials have been limited and have required experimental designs of moderately high complexity. A number of variables can influence the success of this technique including factors related to fixation, tissue pretreatment, efficiency of the PCR reaction, retention of the amplifed product *in situ*, and sensitivity for the detection of the amplified product. IS-PCR should be used only when alternative methods such as IHC, SURF, ISH, or filter hybridization techniques fail to answer the specific scientific or diagnostic question. The development of IS-PCR is time-consuming and technically demanding but it offers great potential for improving our understanding of biological systems.



The enticing promise of IS-PCR is that by amplifying DNA or RNA within cells, the sensitivity of ISH will be elevated to permit the detection of single copy DNA sequences or low copy mRNAs in individual cell preparations or tissue sections. From the time that PCR was first conceived, researchers have recognized

the tremendous potential of being able to combine this technique with ISH. The advantages include elimination of concern over contamination-based false positives, ability to assess the histologic and subcellular distribution of the target, and ability to detect as few as 1 -10 copies in a cell. IS-PCR has been successfully performed using fresh, frozen, cell suspensions, or paraffin-embedded archival fixed tissues.

Methodology

Fixation and pretreatment. Major obstacles have been overcome in the development of IS-PCR. Parameters were established to expose adequately the target nucleic acid sequences in the cells or tissue to the PCR reaction mixture while still maintaining tissue integrity and cellular morphology. Solutions to some of the problems lay in the fixation and appropriate protease treatment of the cells or tissue and the development of equipment designed specifically for IS-PCR.

The fixative of choice is NBF for at least 8 hours(Nuovo, 1994). Choosing a crosslinking fixative is essential because it appears necessary for retention of the amplified products in the tissue. Non-crosslinking acetone or ethanol fixation alone can have deleterious effects on the reaction, and PCR product can easily be washed from the tissue. In general, Carnoy's, Bouin's, and Zenker's solutions should be avoided since they may not achieve sufficient crosslinking, and components of the fixatives may interfere with the PCR (Nuovo, 1994).

Fixation and protease treatment are critical to the success of IS-PCR. The protease treatment must be titrated so that it digests the tissue sufficiently to permit exposure of the target DNA to the reaction mixture, but not so much that the amplified DNA would be washed away. The concentration of protease and the length of time of protease digestion differ according to the specific enzyme used, the fixative, the length of fixation, and the tissue or cell type containing the target.

PCR on slides. There are several other issues that make IS-PCR more difficult than routine solution PCR. To prevent mispriming, the reaction mixture must be added to the tissue at temperatures greater than 70^{o}C, a procedure introduced by Nuovo (1994) called "hot starting". This modification has allowed the detection of as few as one human papilloma virus copy per cell in formalin-fixed, paraffin-embedded samples. Also, the reaction mixture must be sealed over the tissue to prevent evaporation and drying of the tissue during the temperature cycling. Sealing the reaction mixture has been accomplished in a number of ways, including immersing the slide in mineral oil or securing a coverslip over the tissue using nail polish. Some researchers have even suggested mounting tissues onto glass slides cut into pieces small enough to be placed into 0.5 ml PCR tubes containing the reaction mixture.

To achieve adequate thermocycling, initial protocols involved placing the slides on top of thermocycler block designed for solution PCR or in hot air ovens. Although successful results have been achieved, the techniques utilizing investigator-modified solution-PCR equipment may have contributed to inconsistent and irreproducible results. Today, a number of biotechnology companies offer specialized equipment dedicated to IS-PCR that will help optimize thermocyling temperatures while efficiently maintaining reaction mixtures on the slide. This equipment may provide optimal conditions to improve the consistency, reproducibility, and reliability of the technique.

Signal detection. Labeled nucleotide can be incorporated directly into in situ amplificants (direct IS-PCR) or, alternatively, ISH can be performed after in situ amplification using labeled oligonucleotide probes (indirect IS-PCR). For the direct method non-isotopic labels, such as digoxigenin, have been most commonly used. A reportedly high occurrence of false-positive results by the direct method is supposedly due to misprimed DNA products or DNA repair mediated by the *Taq* polymerase. Nuovo (1994) appears to have eliminated some false-positive reactions by pretreating the tissues with a solution of dideoxynucleotides to block DNA repair. Detection by the indirect method is recommended for most experiments and has been performed using either isotopically or non-isotopically labeled probes.

The efficiency of the PCR amplification on the slide is not well defined and has led to some controversy. Some results have suggested that the PCR on the slide is highly inefficient and that perhaps only a 50-fold increase of the original target is produced after 30 cycles (Nuovo, 1991). Despite the inefficient amplification, however, the amplificants are enough to be easily detected by ISH or possibly by a suitably labelled product of direct IS-PCR. Further scientific analysis of the PCR efficiency and specificity may clarify these issues.

In situ RT-PCR. IS-PCR also has been extended to studies of RNA after the in situ synthesis of the corresponding cDNA through the use of reverse transcriptase. This approach has been used successfully for the detection of RNA viruses and for mRNAs. Usually the tissue is first treated with DNase overnight to remove native DNA. Alternatively, it may be possible to design primers that span an intron so that the amplification is specific for the spliced RNA target making DNase treatment unnecessary.

Controls. To confirm the sensitivity and specificity, stringent controls are essential and should include known positive and negative cells or tissues, use of non-complementary primers or oligonucleotides in place of the original primers, amplification mixtures lacking *Taq* or primers, and detection of a known endogenous single-copy gene.

Applications of IS-PCR

IS-PCR has been successful for the detection of low copy number viral DNAs, single-copy genes, immunoglobulin gene rearrangements in cell suspensions and cytospins, and for the analysis of low-level gene expression. The greatest success has been gained in the study of infectious diseases and, in particular, in the diagnosis of viral latency. Applications of IS-PCR for the study of RNA associated with toxicity, carcinogenesis, and cancer progression are being developed.

Potentially, researchers may be able to visualize single cells bearing premalignant mutations or karyotypic alterations. In our laboratory we have tried, unsuccessfully, to identify individual neoplastic and non-neoplastic mouse pulmonary epithelial cells which contain a mutation in the K-*ras* oncogene. Our hope was to elucidate events of tumor initiation and progression using a direct method of IS-PCR on histologic sections of mouse lung. To date, we do not know of any publications in peer-reviewed journals applying IS-PCR for point-mutation analysis. Further refinements of IS-PCR may enhance the utilization of this technique for mutational analysis or perhaps a technique such as laser capture microdissection will suffice.

V. FIXATION

Techniques have been successfully performed using fresh, frozen, fixed, or paraffin-embedded fixed tissues or cell suspensions. Fixation can be detrimental to nucleic acid molecules and thus negatively affect the success of the techniques, however fixation is essential for some of the slide-based techniques (ISH and IS-PCR) and, of course, for the preservation of archival tissues. Fortunately, most of the PCR-based techniques can be successfully applied to paraffin embedded tissues fixed in neutral buffered formalin. There are many optimizing strategies reported to help overcome some of the detrimental effects of fixation on PCR. Shortening the duration of fixation, increasing the initial sample size, amplifying products less than 200 bp, using a nested primer method, and increasing the number of amplifying cycles have all been shown to help achieve positive results for PCR.

For ISH, short fixation of frozen tissue sections with 4% paraformaldehyde seems to achieve the best overall results. Furthermore, when analyzing for RNA, immediate fixation is strongly recommended to inactivate RNA degradation by endogenous ribonucleases. The use of frozen sections for ISH, despite imparting a loss in morphologic detail, obviates the need for deparaffinization and protease digestion -- steps that might contribute to false negative results. For IS-PCR, the fixative of choice is neutral buffered formalin for at least 8 hours. Choosing a crosslinking fixative is essential because it appears necessary for retention of the amplified products in the tissue. Non-crosslinking acetone or ethanol fixation alone can have deleterious effects on the reaction, and PCR product can easily be washed from the tissue. In general, Carnoy's, Bouin's, and Zenker's solutions should be avoided for ISH and IS-PCR since they provide less preservation, they may not achieve sufficient crosslinking, and/or components of the fixatives may interfere with the PCR.

VI. REFERENCES

GENERAL:

Malarkey, DE and Maronpot, RR. (1996). Polymerase chain reaction and *in situ* hybridization: Applications in toxicological pathology. *Tox Path* 24(1): 13-23.

Cotter FE (1994). The role of polymerase chain reaction in pathology. *J. Histotechnol.* 17: 253-259.

Kwok S, and Higuchi R (1989). Avoiding false positives with PCR. *Nature* 339: 237-238.

Grompe M (1993). The rapid detection of unknown mutations in nucleic acids. *Nature Genet.* 5: 111-117.

Loda M (1994). Polymerase chain reaction-based methods for the detection of mutations in oncogenes and tumor suppressor genes. *Hum. Pathol.* 25: 564-571.

PCR/SSCP

Ballhausen WG, and Kraus C (1993). Non-isotopic detection of single-stranded conformation polymorphisms using ethidium bromide/UV light. *Appl. Theor. Electrophor.* 3: 129-131.

Grade K, Grunewald I, Graupner I, Behrens F, and Coutelle C (1994). Identification of three novel mutations in the CFTR gene using temperature-optimized non-radioactive conditions for SSCP analysis. *Human Genetics* 94: 154-158.

Sheffield VC, Beck JS, Kwitek AE, Sandstrom DW, and Stone EM (1993). The sensitivity of single-strand conformation polymorphism analysis for the detection of single base substitutions. *Genomics* 16: 325-332.

PCR/RFLP

Lee, G. and Drinkwater, N.R. 1995. Hepatocarcinogenesis in BXH recombinant inbred strains of mice: Analysis of diverse phenotypic effects of the hepatocarcinogen sensitivity loci. *Molecular Carcinogenesis* 14:190-19

Aguilar F, and Cerutti P (1994). Genotypic mutation analysis by RFLP/PCR. In: *Methods In Toxicology,* CA Tyson, JM Frazier (eds). Academic Press, Inc., San Diego, California, pp. 237-247.

Myers B, and Parry JM (1994). The application of the restriction site mutation assay to the study of the induction of mutations in the tissues of rodents. *Mutagenesis* 9: 175-177.

Comparative genomic hybridization (CGH)

Kallioniemi O, Kallioniemi A, Sudar D, Rutovitz D, Gray JW, Waldman F, and Pinkel D (1993). Comparative genomic hybridization: a rapid new method for detecting and mapping amplification in tumors. *Semin. Cancer Biol.* 4: 41

Breneman JW, Ramsey MJ, Lee DA, Eveleth GG, Minkler JD, and Tucker JD (1993). The development of chromosome-specific composite DNA probes for the mouse and their application to chromosome painting. *Chromosoma* 102: 591-598.

FISH

Wolman SR (1994). Flourescence in situ hybridization: A new tool for the pathologist. *Hum. Pathol.* 25: 586-590.

Celeda D, Bettag U, and Cremer C (1992). A simplified combination of DNA probe preparation and fluorescence *in situ* hybridization. *Z. Naturforsh.* 47C: 739-747.

Loss of heterozygosity (LOH)

Davis LM, Caspary WJ, Sakallah SA, Maronpot RR, Wiseman R, Barrett JC, Elliot R, and Hozier JC (1994). Loss of heterozygosity in spontaneous and chemically induced tumors of the B6C3F1 mouse. *Carcinogenesis* 15: 1637-1645.

Hegi ME, Devereux TR, Dietrich WF, Cochran CJ, Lander ES, Foley JF, Maronpot RR, Anderson MW, and Wiseman RW (1994). Allelotype analysis of mouse lung carcinomas reveals frequent allelic losses on chromosome 4 and an association between allelic imbalances on chromosome 6 and K-*ras* activation. *Cancer Res.* 54: 6257-6264.

Manenti G, DeGregorio L, Gariboldi M, Dragani TA, and Pierotti MA (1995). Analysis of loss of heterozygosity in murine hepatocellular tumors. *Mol. Carcin.* 13: 191-200.

Microsatellite instability

Loeb L (1994). Microsatellite instability: Marker of a mutator phenotype in cancer. *Cancer Res.* 54: 5059-5063.

Shibata, et al. 1996. Somatic microsattelite mutations as molecular tumor clocks. Nature Medicine 2(6):676-681. (Also see related News&views in same issue on p. 630-631)

Strickler, JG et al. 1994 p53 mutations and microsatellite instability in sporadic gastric cancer: when guardians fail. Cancer Research 54:4750-4755.

Shibata, et al. 1994. Genomic instability in repeated sequences is an early somatic event in colorectal tumorigenesis that persists after transformation. Nature Genetics 6 (march 94) 273-281

Canzian F, Ushijima T, Serikawa T, Wakabayashi K, Sugimura T, and Nagao M (1994). Instability of microsatellites in rat colon tumors induced by heterocyclic amines. *Cancer Res.* 54: 6315-6317.

Helicobacter

Ward JM, et al (1994). Chronic active hepatitis and associated liver tumors in mice caused by a persistent bacterial infection with a novel *Helicobacter* species. *Journal of the National Cancer Institute* 86: 1222-1227.

Rice, JM (1995). *Helicobacter hepaticus* , a recently recognized bacterial pathogen, associated with chronic hepatitis and hepatocellular neoplasia in laboratory mice. *Emerging Infectious Diseases* 1: 129-131.

Fox JG, et al (1996). Chronic proliferative hepatitis in A/JCr mice associated with persistent *Helicobacter hepaticus* infection: A model of helicobacter-induced carcinogenesis. *Infection and Immunity* 64: 1548-1558.

Malarkey, DE, et al. (1997). A PCR-RFLP method for the detection of *Helicobacter hepaticus* in frozen or fixed liver from B6C3F1 mice. *Toxicologic Pathology* 25 (6):606-612.

Battles JK, Williamson JC, Pike KM, Gorelick PL, Ward JM, and Gonda MA (1995). Diagnostic assay for *Helicobactor hepaticus* based on nucleotide sequence of its 16S rRNA gene. *Journal of Clinical Microbiology* 33:1344-1347.

Shames B, Fox JG, Dewhirst F, Yan L, Shen Z, and Taylor NS (1995). Identification of widespread *Helicobactor hepaticus* infection in feces in commerical mouse colonies by culture and PCR assay. *Clinical Microbiology* 33:2968-2972.

Riley LK, Franklin CL, Hook RR, and Besch-Williford C (1996). Identification of murine helicobacters by PCR and restriction enzyme analyses. *Clinical Microbiology* 34:942-946.

Beckwith, CS, et al. (1997). Fecal PCR assay for diagnosis of *Helicobacter* Infection in laboratory rodents. *Journal of Clinical Microbiology* 35:1620-1623.

Mahler, M, et al. (1998). Comparison of four diagnostic methods for detection of *Helicobacter* species in laboratory mice. *Laboratory Animal Science* 48:85-91.

Fox, JG, et al. (1998). Comparison of methods of identifying *Helicobacter hepaticus* in B6C3F1 mice used in a carcinogenesis bioassay. *Journal of Clinical Microbiology* 36: 1382-1387.

Livingston, RS, et al. (1997). Serodiagnosis of *Helicobacter hepaticus* Infection in Mice by an Enzyme-Linked Immunosorbent Assay. *Journal of Clinical Microbiology* 35: 1236-1238.

ISH

Brown, C. (1998). In Situ Hibridization with Riboprobes: An Overview for Veterinary Pathologists. *Veterinary Pathology* 35: 159-167.

DeLellis RA (1994). In situ hybridization techniques for the analysis of gene expression: applications in tumor pathology. *Hum. Pathol.* 25: 580-585.

Montone KT, and Brigati DJ (1994). In situ molecular pathology: instrumentation, oligonucleotides, and viral nucleic acid detection. *J. Histotechnol.* 17: 195-201.

Wilcox JN (1993). Fundamental principles of in situ hybridization. *J. Histochem. Cytochem.* 41: 1725-1733

Herrington CS, and McGee JO (1992). Principles and basic methodology of DNA/RNA detection by *in situ* hybridization. In: *Diagnostic Molecular Pathology*, CS Herrington, JO McGee (eds). Oxford University Press, New York, pp. 69-102.

Alison M, Chaudry Z, Baker J, and Lauder I (1994). Liver regeneration: A comparison of in situ hybridization for histone mRNA with bromodeoxyuridine labeling for the detection of S-phase cells. *J. Histochem. Cytochem.* 42: 1603-1608

Baskin DG, and Stahl WL (1993). Fundamentals of quantitative autoradiography by computer densitometry for in situ hybridization with emphasis on ^{33}P. *J. Histochem. Cytochem.* 41: 1767-1776.

Chou MY, Chang ALC, McBride J, Donoff B, Gallagher GT, and Wong DTW (1990). A rapid method to determine proliferation patterns of normal and malignant tissues by H3 mRNA *in situ* hybridization. *Am. J. Pathol.* 136: 729-733.

differential display

Bauer D, Muller H, Reich J, Riedel H, Ahrenkiel V, Warthoe P, and Strauss M (1993). Identification of differentially expressed mRNA species by an improved display technique (DDRT-PCR). *Nucleic Acids Res.* 21: 4272-4280.

Douglass J, McKinzie AA, and Couceyro P (1995). PCR differential display identifies a rat brain mRNA that is transcriptionally regulated by cocaine and amphetamine. *J. Neurosci.* 15: 2471-2481.

Liang P, Averboukh L, Keyomarsi K, Sager R, and Pardee AB (1992). Differential display and cloning of messenger RNAs from human breast cancer versus mammary epithelial cells. *Cancer Res.* 52: 6966-6968.

Liang P, Averboukh L, and Pardee AB (1993). Distribution and cloning of eukaryotic mRNAs by means of differential display: refinements and optimization. *Nucleic Acids Res.* 21: 3269-3275.

Liang P, and Pardee AB (1992). Differential display of eukaryotic messenger RNA by means of the polymerase chain reaction. *Science* 257: 967-971.

Wang X, and Feuerstein GZ (1994). Direct sequencing of DNA isolated from mRNA differential display. *Biotechniques* 18: 448-452.

quantitative rtPCR (QRT-PCR)

Freeman, WM, Walker, J, and Vrana, KE (1999) Quantitative RT-PCR: Pitfalls and potential. Biotechniques 26:112-125.

Vanden Heuvel JP, Clark GC, Kohn MC, Tritscher AM, Greenlee WF, Lucier GW, and Bell DA (1994). Dioxin-responsive genes: examination of dose-response relationships using quantitative reverse transcriptase-polymerase chain reaction. *Cancer Res.* 54: 62-68.

Vanden Heuvel JP, Clark GC, Thompson CL, McCoy Z, Miller CR, Lucier GW, and Bell DA (1993). CYP1A1 mRNA levels as a human exposure biomarker: use of quantitative polymerase chain reaction to measure CYP1A1 expression in human peripheral blood lymphocytes. *Carcinogenesis* 14: 2003-2006.

Vanden Heuvel JP, Tyson FL, and Bell DA (1993). Construction of recombinant RNA templates for use as internal standards in quantitative RT-PCR. *Biotechniques* 14: 395-398

Riedy MC, Timm J E.A., and Stewart CC (1995). Quantitative RT-PCR for measuring gene expression. *Biotechniques* 18: 70-76.

Foley KP, Leonard MW, and Engel JD (1993). Quantitation of RNA using the polymerase chain reaction. *Trends Genet.* 9: 380-385.

Gilliland G, Perrin S, and Bunn HF (1990). Competitive PCR for quantitation of mRNA. In: *PCR Protocols: A guide to Methods and Applications,* M Innis, D Gelfand, J Sninsky, T White (eds). Academic Press, San Diego, California, pp. 60-69.

Riedy MC, Timm J E.A., and Stewart CC (1995). Quantitative RT-PCR for measuring gene expression. *Biotechniques* 18: 70-76.

microarray / DNA chip technology

Schena, M, et al. (1995). Quantitative monitoring of gene expression patterns with a complementary DNA microarray. *Science* 270: 467-470.

DeRisi, J, et al. (1996). Use of a cDNA microarray to analyxe gene expression patterns in human cancer. *Nature Genetics* 14: 457-460.

Heller, RA, et al. (1997). Discovery and analysis of inflammatory disease-related genes using cDNA microarrays. *Proc. Natl. Acad. Sci. USA* 94: 2150-2155.

Shalon, D, et al. (1996). A DNA microarray system for analyzing complex DNA samples using two-color fluorescent probe hybridization. *Genome Research* 6: 639-645.

Ramsay, G. (1997). DNA chips: state-of-the art. *Nature Biotechnology* 16: 40-44.

Pardee, AB. (1997). Complete genome expression monitoring: the human race. *Nature Biotechnology* 15: 1343-1344.

Schena, M, et al. (1996). Parallel human genome analysis: microarray-based expression monitoring of 1000 genes. *Proc. Natl. Acad. Sci. USA* 93: 10614-10619.

Blanchard, AP and Leroy Hood. (1996). Sequence to array: probing the genome's secrets. *Nature Biotechnology* 14: 1649.

PNA (peptide nucleic acids)

Wang, J (1998). DNA biosensors based on PNA recognition layers. A review. Biosensors and bioelectronics 13:757-762.

Nielsen, PE. (1999). Applications of peptide nucleic acids. Current Opinion in Biotechnology 10:71-75.

Nielsen, PE, Egholm, M, Berg, RH, and Burchardt, O. (1991). Sequence selective recognition of DNA by strand displacement with a thymine-substituted polyamide. Science 254:1497-1500.

Egholm, M Buchardt, O, et al. (1993). PNA hybridizes to complementary oligonucleotides obeying the Watson-Crick hydrogen bonding rules. Nature 365:556-568.

Ortiz, E. Estrada, G. and Lizardi, PM. (1998). PNA molecular beacons for rapid detection of PCR amplicons. Molecular and Cellular Probes 12:219-226.

Microdissection of tissue

d'Amore, F, Stribley, JA, Ohno, T, Wu, G, Wickert, RS, Delabie, J, Hinrichs, SH, and Chan, WC. 1997. Molecular studies on single cells harvested by micromanipulation from archival tissue sections previously stained by immunohistochemistry or nonisotopic in situ hybridization. *Lab Investigation* 76(2):219-224.

Turbett, GR, Barnett, TC, Dillon, EK, and Sellner, LN. (1996). Single-tube protocol for the extraction of DNA or RNA from paraffin-embedded tissues using a starch-based adhesive. *Biotechniques* 20(5): 846-851

SURF

Shibata D, Hawes D, Li Z-H, Hernandez AM, Spruck CH, and Nichols PW (1992). Specific genetic analysis of microscopic tissue after selective ultraviolet radiation fractionation and the polymerase chain reaction. *Am. J. Pathol.* 141: 539-543.

Li Z-H, Zheng J, Weiss LM, and Shibata D (1994). c-K-ras and p53 mutations occur very early in adenocarcinoma of the lung. Am. J. Pathol 144: 303-309.

Strickler, JG et al. 1994 p53 mutations and microsatellite instability in sporadic gastric cancer: when guardians fail. Cancer Research 54:4750-4755.

Shibata, et al. 1994. Genomic instability in repeated sequences is an early somatic event in colorectal tumorigenesis that persists after transformation. Nature Genetics 6 (march 94) 273-281

Zheng, et al. 1994. Patterns of p53 mutations in SCC of the lung: Acquisition at a relatively early age. Am J Path 145(6): 1444-1449.

Laser capture microdissection

Emmert-Buck, MR, et al. 1996. Laser Capture Microdissection. Science 274 (nov 8) 998-1001.

Bonner, RF, Emmert-Buck MR, Cole, K, Pohida, T, Chuaqui, RF, Goldstein, SR, and Liotta, LA. 1997. Laser Capture Microdissection:Molecular analysis of tissue. *Science* 278: 1481-1483.

IN SITU PCR (IS-PCR)

Nuovo GJ (1994). *PCR in situ Hybridization: Protocols and applications*. (2nd ed.) Raven Press, Ltd, New York.

Bagasra O, Seshamma T, Hansen J, Bobroski L, Saikumari P, Pestaner JP, and Pomerantz RJ (1994). Application of *in situ* PCR methods in molecular biology: I. Details of methodology for general use. *Cell Vision* 1: 324-335.

Long AA, Komminoth P, Lee E, and Wolfe HJ (1993). Comparison of indirect and direct in-situ polymerase chain reaction in cell preparations and tissue sections. *Histochemistry* 99: 151-162.

Heniford BW, Shum-Siu A, Leonberger M, and Hendler FJ (1993). Variation in cellular EGF receptor mRNA expression demonstrated by *in situ* reverse transcriptase polymerase chain reaction. *Nucleic Acids Res.* 21: 3159-3166.

Komminoth P, and Long AA (1994). Correspondence: Questioning *in situ* PCR. *Am. J. Pathol.* 145: 742.

Komminoth P, Long AA, Ray R, and Wolfe HJ (1992). In situ polymerase chain reaction detection of viral DNA, single-copy genes, and gene rearrangements in cell suspensions and cytospins. *Diag. Mol. Pathol.* 1: 85-

Fixation

Shibata, D. (1994). Extraction of DNA from paraffin-embedded tissue for analysis by polymerase chain reaction: new tricks from an old friend. Human Pathology 25(6):561-563.

Mies C (1994). Molecular biological analysis of paraffin-embedded tissues. *Hum. Pathol.* 25: 555-560.

Ben-Ezra J, Johnson D, Rossi J, Cook N, and Wu A (1991). Effect of fixation on the ampification of nucleic acids from paraffin embedded material by the polymerase chain reation. Histochemistry and cytochemistry 39: 351-354.

Thompson CH, and Rose BR (1991). Deleterious effects of formalin/acetic acid/alcohol (FAA) fixation on the detection of HPV DNA by in situ hybridization and the polymerase chain reaction. *Pathology* 23: 327-330.

Gianella, C, et al. (1997). Comparison of formalin, ethanol, and histochoice fixation on the PCR amplification from paraffin-embedded breast cancer tissue. *European Journal of Clinical Chemistry and Clinical Biochemistry* 35: 633-635.

Hamazaki, S, et al. (1993). The effect of formalin fixation on restriction endonuclease digestion of DNA and PCR amplification. *Path.Res.Pract.* 189: 553-557.

Crisan, D, et al. (1990). Polmerase chain reaction: amplification of DNA from fixed tissue. *Clinical Biochemistry* 23: 489-495.

Rogers, BB, et al. (1990). Analysis of DNA in fresh and fixed tissue by the polymerase chain reaction. *American Journal of Pathology* 136: 541-548.

GLOSSARY

A: Abbreviated symbol for the base adenine.

Allele: One of several alternate forms of a gene which occupies a specific locus on a chromosome.

Allele: One of the two gene pairs situated at the same location (locus) on a chromosome; one allele is inherited from the mother and the other from the father.

Allelic exclusion: The expression in a particular lymphocyte of one specific allele which codes for the expressed immunoglobulin.

Alu family: A set of dispersed, related sequences in the human genome. Each is approximately 300 bp long and can be cleaved at each end by an Alu restriction endonuclease.

Amber codon: The nucleotide triplet UAG, a "nonsense" codon that causes termination of protein synthesis.

Amplification: The production of added copies of a chromosomal sequence.

Annealing: The pairing of complementary single strands of nucleic acid**.**

Anticodon: A trinucleotide sequence on a tRNA molecule which is complementary to the codon representing its amino acid, enabling the tRNA to recognize the codon on the mRNA via complementary base pairing.

Antiparallel: Refers to the strands of duplex nucleic acid which are organized in opposite orientation, with the 5' end of one strand aligned with the 3' end of the other strand.

Base pair: A partnership of A with T or of C with G in a double stranded DNA. Often abbreviated "bp."

Biallelic damage: Damage to both maternal and paternal copies of a gene.

Bioinformatics: The computational biological process designed to handle and interpret the streams of DNA data entering various computer data bases in association with the effort to sequence the human genome

Blunt-end: Refers to two strands of DNA whose ends are of identical length, with complete base pairing. See staggered cuts or skickey ends.

bp: Abbreviation for "base pairs." Often used as a measure for the distance along DNA.

C: Abbreviation for the base cytosine.

CAAT box: A conserved sequence, located approximately 75 bp upstream of the

startpoint of most eukaryotic transcription units. It is recognized by a large group of transcription factors.

Cap: The structure at the 5' end of eukaryotic mRNA, which is added after transcription by linking the terminal phosphate of 5' GTP to the terminal base of the mRNA.

cDNA clone: A double stranded DNA sequence which represents an RNA, carried in a cloning vector.

cDNA: Abbreviation for "complementary DNA," a single stranded DNA molecule which is complementary to an RNA. It is synthesized in vitro from the RNA by reverse transcription.

CGAP: Cancer Genome Anatomy Project; objective is to identify all human cancer genes. There is also a mouse CGAP.

Chromatin: The complex of DNA and protein in the nucleus of the cell.

Chromosome walking: The process by which clones carrying overlapping sequences of DNA are sequentially isolated, allowing large regions of the chromosome to be spanned.

Cis-acting locus: A property which affects the activity only of DNA sequences on its own molecule of DNA, usually implying that the locus is in the DNA sequence itself and does not code for protein.

Cistron: A segment of DNA or RNA that encodes a single polypeptide chain.

Clonal: pertaining to a clone; a line of cells decendent from a single cell.

Clone: A large number of cells or molecules identical to a single ancestral cell or molecule.

Cloninq vector: A plasmid or a phage which is used to carry inserted foreign DNA.

Closed reading frame: Refers to a reading frame of a nucleotide sequence which contains termination codons which would prevent translation into protein.

Coding strand: A strand of DNA which has the same sequence as mRNA. Often called the positive strand. The complementary or negative strand is referred to as the template strand.

Codon: A triplet of nucleotides which represents an amino acid or a termination (nonsense) signal.

Concatemer: A series of unit genomes which are repeated in tandem.

Consensus sequence: An idealized sequence in which each position represents the base most often found when many actual sequences are compared.

Constitutive genes: Genes which are expressed all the time, without added regulatory processes. Also called housekeeping genes.

Cotransfection: The simultaneous transfection of two markers.

Cruciform: A structure produced at inverted repeats of DNA when the repeated sequence pairs with its complement on the same strand.

Degeneracy: Refers to situations where there is lack of an effect of the third base of the codon on the amino acid that is represented.

Deletion: The removal of a sequence of DNA with the regions on either side being joined together.

Denaturation: The conversion of DNA from double stranded to single stranded state.

Direct repeats: Identical or related sequences which are present in two or more copies in the same orientation in the same molecule of DNA. The repeats do not have to be adjacent.

DNA: abbreviation for deoxyribonucleic acid; the basic building block of genetic material in all organisms except RNA viruses.

DNA: abbreviation for deoxyribonucleic acid; the basic building block of genetic material in all organisms except RNA viruses

DNAse: An enzyme that attacks bonds in DNA.

DNA polymerase: An enzyme that synthesizes a daughter strand of DNA in either replication or repair.

DNA replicase: An enzyme that synthesizes DNA and is specifically required for replication.

Domain: A discrete structural antity within a chromosome where supercoiling is independent of other domains. May also refer to an extensive region including an expressed gene with heightened sensitivity to degradation by DNAase.

Downstream: Sequences proceeding farther in the direction of expression, i.e. toward the 3' end of the nucleic acid.

Elongation factors: Proteins which associate with ribosomes in a cyclic fashion for the sequential addition of each amino acid to the polypeptide chain.

End labeling: The addition of a radioactive or fluorescent labeled group to one end (either 5' or 3') of a nucleic acid.

Endonuclease: An enzyme which cleaves bonds within a nucleic acid chain. It may be specific for RNA, DNA, single stranded DNA, or double stranded DNA.

Enhancer: An element which is cis-acting, which increases the utilization of some eukaryotic promoters, and which can function in either orientation and in any location (downstream or upstream) relative to the promoter.

Epigenetic: Changes which influence the phenotype without changing the genotype.

EST: Expressed sequence tag. Short coding sequences (~150-300 nucleotides) representing portions of expressed genes. A current effort is underway to catalog thousands of EST's and identify the genes from which they originate.

Eukaryote: An organism whose genes are located in chromosomes inside a nucleus.

Exon: A segment of an interrupted gene that is represented in the mature RNA product.

Exonuclease: An enzyme which cleaves nucleotides one at a time from the end of a polynucleotide chain. The enzyme may be specific for either DNA, RNA, or the 3' or 5' end.

Expression vector: A cloning vector which is designed so that a coding sequence inserted at a particular site will be transcribed and translated into protein.

Filter hybridization: The incubation of a denatured nucleic acid (usually DNA) preparation immobilized on a filter with a solution of radioactively or fluorescently labeled RNA or DNA.

Fingerprint: A pattern within DNA where there is a pattern of polymorphic restriction fragments that differ between individual genomes.

Five Prime (5'): Refers to the 5' carbon position of ribose or deoxyribose to which phosphate groups may be attached; signifying one end of a polynucleotide chain.

Footprinting: A technique to identify the site on DNA which is bound by a protein by virtue of the protection of phosphodiester bonds in this region against attack by nucleases.

Frameshift mutation: A mutation which arises by deletions or insertions which are not a multiple of 3 bp. This results in a change in the reading frame in which triplets are translated into protein.

G: Abbrevation of the base guanine.

Gap: The absence of one or more nucleotides in one of the strands of duplex DNA.

Gene cluster: A group of adjacent genes which are identical or related.

Gene dosage: The number of copies of a particular gene in the genome.

Gene family: A set of genes whose exons are related, usually derived by duplication and variation from an ancestral gene.

Gene: A segment of DNA involved in producing a polypeptide chain or a specific RNA molecule. It is sometimes called a cistron. A gene may include regions preceding and following the coding regions (leader, trailer) as well as introns between individual exons.

Gene: the basic biological unit of heredity which is located on a chromosome.

Gene: the basic biological unit of heredity which is located on a chromosome.

Genetic code: The correspondence between triplets in DNA or RNA and the amino acids in protein.

Genome: the total complement of genes present in the set of chromosomes characteristic of a given organism.

Genome: the total complement of genes present in the set of chromosomes characteristic of a given organism.

Genomic clone: A clone which contains sequences of the genome carried by a cloning vector.

Genotoxic: toxic to DNA; an agent or process that interacts with cellular DNA either directly or after metabolic transformation; mutagens are genotoxic agents.

Genotoxic: toxic to DNA; an agent or process that interacts with cellular DNA either directly or after metabolic transformation; mutagens are genotoxic agents.

Genotype: The genetic constitution of an organism.

Growth factors: agents that contribute to and stimulate tissue growth.

GT-AG rule: The presence of specific dinucleotides at the first two and the last two positions of introns of nuclear genes.

Hairpin: A double helical region formed by base pairing between adjacent, inverted, complementary sequences in a single strand of nucleic acid.

Haploid: Refers to the set of chromosomes which contains one copy of each autosome and one sex chromosome.

Hemizygote: A diploid individual that has lost one of the two copies of a particular gene (e.g., because a chromosome or part of a chromosome has been lost), resulting in only a single copy of that gene. A transgenic animal may be hemizygous for a particular transgene if that animal has only one copy of the transgene.

Heterochromatin: Regions of the genome which are permanently in a highly condensed condition and are not genetically expressed.

Heteroduplex DNA: DNA which is generated by base pairing between complementary single strands derived from the different parental duplex molecules. Also called hybrid DNA.

Heterozygote: An individual with different alleles at a particular 1ocus .

Heterozygous: having different alleles at a specific position on a chromosome.

Heterozygous: having different alleles at a specific position on a chromosome.

Histones: Conserved DNA binding proteins of eukaryotes which form the nucleosome, the basic subunit of chromatin.

Hogness box: Also called the TATA box. A conserved A-T rich septamer which is approximately 25 bp upstream of the startpoint of nuclear genes and which is thought to be involved in the correct positioning of RNA polymerase II for initiation of transcription.

Homeo box: The conserved sequence which is a part of the coding region of Drosophila melanogaster homeotic genes. It is also found in amphibian and mammalian genes expressed in early embryonic development.

Homeotic genes: Genes which are defined by mutations that convert one body part into another in Drosophila melanogaster.

Homologous: Chromosomes which carry the same genetic loci. A diploid cell therefore has two copies of each homologue, one from each parent.

Homozygote: An individual with the same allele at a particular locus on the homologous chromosomes.

Homozygous: having identical alleles at a specific position on a chromosome.

Hotspot: A site at which the frequency of mutation or recombination is much increased.

Housekeeping genes: Genes theoretically expressed in all cells because they provide for basic functions needed for sustenance of all cell types. (See constitutive).

Hybridization: The pairing of complementary RNA or DNA strands to give a double stranded nucleic acid.

Hyperchromatocity: The increase in optical density that occurs when DNA is denatured.

Hypersensitive site: A short region of chromatin which is detected by its extreme sensitivity to cleavage by DNAse I or other nucleases. It probably comprises an area from which nucleosomes are excluded.

Imprinting: Describes a change in a gene that occurs during passage through the sperm or egg and results in the paternal and maternal alleles having different properties of expression, particularly important in early embryogenesis and appearing to play a role in some cancers. may be caused by methylation of DNA.

In situ hybridization: The technique of hybridization applied to cells to provide the precise localization of a specific nucleic acid sequence.

Inducer: A molecule that triggers gene transcription by binding to a regulatory protein.

Induction: The switching on of transcription as a result of interaction of an inducer with a regulatory protein, resulting in increased gene expression.

Initiation factors: Proteins which associate with the small subunit of the ribosome specifically at the stage of initiation of protein synthesis.

Insertion: The presence of additional base pairs of DNA.

Intercistronic region: The distance between the termination codon of one gene and the initation codon of the next gene.

Intervening sequence: **See intron.

Intron: A segment of DNA which is transcribed, but is later removed from within the transcript by splicing together the sequences (exons) on either side of it. Also called intervening sequence.

Inverted repeat: Two copies of the same sequence of DNA which are repeated in opposite orientation to each other on the same molecule. If the inverted repeats are adjacent to each other, a palindrome is formed.

Karyotype: The entire chromosomal complement of a cell visualized during mitosis.

kb: Abbreviation for 1000 base pairs of DNA or 1000 bases of RNA.

Kinase: An enzyme that adds a gamma phosphate from ATP to the 5' end of a polynucleotide chain.

Knockout: An experiment in which a defective gene replaces a normal gene in the genome of an organism.

Leader: The nontranslated sequence at the 5' end of mRNA which precedes the initiating methionine.

Library: A set of cloned fragments which together represent either an entire genome or a representative sample of a particular subset of expressed genes.

Ligation: The formation of a phosphodiester bond to link two bases.

Linkage: The tendency of genes to be inherited together as a result of their location on the same chromosome.

Linker fragment: A short synthetic duplex oligonucleotide which contains the target site for a restriction enzyme, which may be added to ends of a DNA fragment during reconstructions of recombinant DNA.

Locus: The position on a chromosome at which the gene for a particular trait is located.

LTR: Abbreviation of "long terminal repeat," a sequence directly repeated at both ends of a retroviral DNA.

Luxury genes: Genes coding for specialized functions and generally produced in large amounts in particular cell types.

Map distance: A measurement of the distance between points on a chromosome, measured as centiMorgans (cM), as determined by the percent recombination between two loci.

Marker: A fragment of known size used to calibrate an electrophoretic gel.

Melting temperature: The midpoint of the temperature range over which DNA is denatured, often abbreviated as Tm.

Melting: The denaturation of DNA.

Microsatellites (short tandem repeats): Dinucleotide (and sometimes trinucleotide) repeats (e.g., CA:GT) scattered throughout the genome. The number of repeats at any given locus varies between individuals and, thus, may be used as a genetic marker. A microsatellite instabity assay has been used as a generic measure of genetic damage.

Modification: Changes made to the nucleotides after their initial incorporation into the polynucleotide chain.

Modified base: A base other than the usual T, C, A, G, or U which results from postsynthetic changes in the nucleic acid.

mRNA: Abbreviation of messenger RNA.

Mutation: A change in the sequence of genomic DNA.

Mutation: a structural alteration in DNA that is hereditary and may give rise to an altered phenotype.

Mutation: a structural alteration in DNA that is hereditary and may give rise to an altered phenotype.

Negative strand: Sometimes used to refer to that strand of duplex DNA which is the template to direct the synthesis of RNA that is complementary to it; the negative strand is complementary to the coding strand.

Nick translation: A technique used to introduce labeled nucleotides into DNA in vitro, made possible by the ability of E. coli DNA polymerase I to use a nick as a starting point from which one strand of a duplex DNA is degraded and then replaced by the resynthesis of new material containing the label.

Nick: The absence of a phosphodiester bond between two adjacent nucleotides on one strand of a duplex DNA.

Nonsense codon: Any one of three triplets (UAG, UAA, UGA) that cause termination of protein synthesis. Also called amber (UAG), ochre (UAA), umber (UGA).

Nonsense mutation: A change in the DNA that causes a nonsense codon to replace a codon representing an amino acid.

Northern blotting: A technique for transferring RNA from an agarose gel to a filter on which it can be hybridized to complementary DNA.

Nucleosome: The basic structural subunit of chromatin, consisting of approimately 200 bp of DNA and an octamer of histone proteins.

Nucleotide: a biochemical component of DNA that consists of a purine or pyrimidine base, a ribose or deoxyribose sugar, and a phosphate group; a basic building block of DNA.

Nucleotide: a biochemical component of DNA that consists of a purine or pyrimidine base, a ribose or deoxyribose sugar, and a phosphate group; a basic building block of DNA.

Ochre codon: The triplet UAA, one of three nonsense codons that causes termination of protein synthesis.

Oligonucleotide: A short stretch of DNA (or RNA) consisting of just a few nucleotides.

Oligo-dT: A short chain of DNA whose bases are all thymine.

Oncogene activation: the process whereby a protooncogene is altered such that it stimulates enhanced cellular growth; several different mechanisms can lead to such activation.

Oncogene activation: the process whereby a protooncogene is altered such that it stimulates enhanced cellular growth; several different mechanisms can lead to such activation.

Oncogene: a 'so-called' cancer gene because alterations in its structure or expression are typically associated with neoplasms; an activated form of a protooncogene.

Oncogene: a 'so-called' cancer gene because alterations in its structure or expression are typically associated with neoplasms; an activated form of a protooncogene.

Oncogenes: Genes whose products have the ability to transform eukaryotic cells so that they grow like tumor cells. They are carried by retroviruses.

Open reading frame: Refers to a series of triplets coding for amino acids resulting in a sequence which is potentially translatable into protein.

Operator: The site on DNA at which a repressor protein binds to prevent transcription from initiating at the adjacent promoter.

Operon: A unit of bacterial gene expression and regulation, often containing a series of related structural genes and DNA control elements recognized by regulatory proteins. The cluster of genes is often transcribed together to give a single molecule of mRNA.

Origin (ori): The sequence of DNA at which replication is initiated.

Palindrome: A sequence of DNA that is the same when one strand is read left to right, or the other strand is read right to left, consisting of adjacent inverted repeats.

PCR: Polymerase chain reaction

Periodicity: Refers to the number of bp per turn of the double helix.

Phenotype: The characteristics (outward appearance) of an organism which result from the interaction of the genetic constitution and the environment.

Phenotype: the physical appearance, biochemical makeup, and physiological behavior of an individual.

Phosphatase: An enzyme that removes the 5' phosphate from a polynucleotide chain.

Plasmid: An autonomous self-replicating extrachromosomal cirular DNA.

Ploidy: The number of copies of the chromosome set which is present in a cell**:** e.g., haploid has one copy, diploid has two copies, and polyploid has more than two copies, etc.

Point mutation: A change involving a single base pair.

PolyA tail: A string of adenines on the 3' end of mRNA in eukaryotes.

Polyadenylation: The addition of a sequence of polyadenylic acid (multiple A's) to the 3' end of a eukaryotic RNA after transcription.

Polymerase chain reaction: The enzymatic synthesis and amplification of specific DNA sequences. Referred to as PCR.

Polymorphism: The simultaneous occurrence in the population of genomes showing allelic variations.

Polysome: A polyribosome, with a mRNA associated with a series of ribosomes engaged in translation.

Positive strand: Sometimes used to refer to the coding strand of duplex DNA, having the same sequence as the mRNA.

Primary transcript: The original and unmodified RNA product corresponding to a transcription unit. In eukaryotic cells, it contains the introns.

Primer: A short sequence which is paired with one strand of DNA to provide a free 3' OH end at which a DNA polymerase starts synthesis of a deoxyribonucleotide chain. A short sequence of nucleic acid that binds to the template strand and allows synthesis of a new chain of DNA to commence. RNA primers are used in vivo and DNA primers are used in the PCR reaction.

Probe: A nucleic acid which is labeled (usually radioactively or with a fluorescent dye) and can hybridize to another nucleic acid on the basis of complementarity.

Promoter: A region of DNA involved in binding of RNA polymerase to initiate transcription.

Proto-oncogenes: The normal counterparts in the eukaryotic genome to the oncogenes carried by retroviruses, usually given names of the form c-onc; normal cellular structural genes that, when activated by mutations, amplifications, rearrangements, or viral transduction, functions as oncogenes and are associated with neoplasia; regulate normal processes related to cell growth and differentiation.

Pseudogene: An inactive but stable component of the genome which is derived by mutation of an ancestral active gene.

Purine: Fused five and six member rings. The bases adenine (A) and guanine (G) are purines**.**

Pyrimidine: Six member rings. The bases thymine (T), uracil (U), and cytosine (C) are pyrimidines.

Reading frame: **One of three possible ways of reading a nucleotide sequence as a series of triplets.

Reassociation: The pairing of complementary single strands of DNA to form a double helix.

Recognition site: Specific base sequence where a restriction enzyme binds.

Regulatory gene: A gene which encodes an RNA or protein product whose function is to control the expression of other genes.

Renaturation: The reassociation of denatured complementary single strands of a DNA double helix. May also be used to descibe the recovery of structure of a denatured protein.

Repression: Inhibition of transcription or translation by the binding of a repressor protein to a specific site on DNA or RNA.

Restriction enzyme: An enzyme (endonuclease) which recognizes a specific short sequence of DNA and cleaves the DNA either at that site or at some defined distance from it.

Restriction fragment length polymorphism: Refers to a change in the length of a restriction fragment due to an inherited difference in sites for restriction enzymes. Abbreviated form is RFLP.

Restriction fragment: Refers to a fragment of DNA which has been cleaved from a larger piece of DNA by the action of a specific restriction enzyme.

Restriction map: A linear array of sites on DNA cleaved by various restriction enzymes.

Retrovirus: An RNA virus that propagates via conversion into duplex DNA.

Reverse transcription: The synthesis of DNA on a template of RNA, accomplished by reverse transcriptase enzyme.

Reverse translation: A technique for isolating genes or cDNAs by their ability to hybridize with a short oligonucleotide sequence which is prepared by predicting the nucleic acid sequence from the known protein sequence.

RFLP: Restriction fragment length polymorphism

RNAse: An enzyme which attacks bonds in RNA.

rRNA: Abbreviation of ribosomal RNA.

S1 nuclease: An enzyme that specifically degrades single stranded sequences of DNA.

Saturation hybridization: A hybridization experiment in which a large excess of one component is present, causing all complementary sequences in the other component to enter a duplex form.

Selection: The use of particular conditions, e.g. antibiotic resistance, to allow survival only of cells with a particular phenotype.
Shotgun experiment: The cloning of an entire genome or set of DNA fragments in the form of randomly generated fragments.

Shuttle vector: A vector, often a plasmid, which is constructed to have origins for replication for two hosts, so that it can be used to carry a foreign sequence in either prokaryotes or eukaryotes.

Silent mutation: A change in the DNA which does not change the product of the gene.

snRNA: Small nulcear RNA species which are in the nucleus and which may be involved in splicing or other RNA processing reactions.

Snurps: Small nulcear ribonucleoprotiens; snRNAs associated with proteins.

Somatic cells: Cells of an organism which are not those of the germ line.

Somatic mutation: A mutation occurring in a somatic cell, affecting only the descendants of that particular cell, and therefore not inherited by the individual.

Southern blottinq: The procedure by which denatured DNA is transferred from an agarose gel to a filter on which it can be hybridized with a complementary nucleic acid.

Splicing junctions: The sequences immediately surrounding the exon-intron boundaries, which often show conserved features.

Splicing: The removal of introns and the joining of exons in RNA.

Spontaneous mutation: A mutation which occurs in the absence of any known added reagent to increase the mutation rate.

ssDNA: Single stranded DNA.

Staggered cuts: The cuts which are made in duplex DNA when two strands are cleaved (usually by a restriction enzyme) at different points near each other. The result is a duplex DNA with either an overhang of 5' or 3' sequences on each strand, called sticky ends.

Stem: The base paired segment of a hairpin.

Sticky ends: Complementary single strands of DNA that protrude from opposite ends of a duplex molecule, often generated by staggered cuts in duplex DNA.

Stop codon: Codon that signals the end of a protein. Also called nonsense codon.

Structural gene: A gene which codes for any RNA or protein product other than a regulator.

STS: A sequence tag site that is used in the human genome project to construct overlapping maps of gene fragments.

Subcloning: The process in which a fragment of a gene that has already been cloned is recloned into a vector.

Suppression: The occurrence of changes that eliminate the effects of a mutation, without reversing the original change in the DNA.

Suppressor: A compensating mutation that restores the original meaning of a gene. It may be extragenic, undoing a mutant tRNA which reads the mutated codon in the sense of the original codon, or it may be intragenic, restoring the original reading frame after a frameshift mutation.

T: Abbreviation for the base thymine.

Tandem repeats: Multiple copies of the same sequence which lie in a series.

TATA box: Also called a Hogness box. A conserved A-T rich septamer found approximately 25 bp before the startpoint of most eukaryotic nuclear genes and thought to be involved in positioning the RNA polymerase II for correct initiation of transcription.

Termination codon: Also called a nonsense codon. One of three triplet sequences, UAG, UAA, or UGA that cause termination of protein synthesis.

Thymine dimer: A chemically cross-linked pair of adjacent thymine residues in DNA, the result of damage induced by ultraviolet radiation.

T_m: Abbreviation for melting temperature.

Trailer: The nontranslated sequence at the 3' end of an mRNA following the termination codon.

Trans: Refers to the presence of two sites on two different molecules of DNA.

Transcription: The synthesis of RNA on a DNA template.

Transduction: The transfer of a bacterial gene from one bacterium to another by a phage.

Transfection: The acquisition of new genetic material by incorporation of added DNA into eukaryotic cells.

Transformation: In bacteria, the acquisition of new genetic markers by incorporation of added DNA, usually by a plasmid. In eukaryotic cells, refers to the conversion to a state of unrestrained growth in culture, resembling the tumorigenic condition.

Transgenic mouse: A mouse in which a foreign gene has been introduced into the germline, with propagation of the added genetic material to the progeny.

Transition: A mutation in which one pyrimidine is substituted by the other, or in which one purine is substituted for the other.

Translation: The synthesis of protein on the mRNA template.

Transition: A mutation in which one pyrimidine is substitute for another pyrimidine or one purine is substitute for another purine.

Translocation: A rearrangement in which one part of a chromosome is detached by breakage and then becomes attached to another chromosome.

Transversion: A mutation in which a purine is replaced by a pyrimidine, or vice versa.

tRNA: Abbreviation of transfer RNA.

Tumor suppressor gene: a gene that normally functions to suppress uncontrolled tissue growth by inhibiting the activity of oncogenes; sometimes called an ‘ anti-oncogene.

Tumor suppressor gene: a gene that normally functions to suppress uncontrolled tissue growth by inhibiting the activity of oncogenes; sometimes called an ‘ anti-oncogene.’

U: Abbreviation of the base uracil.

Umber codon: The nucleotide triplet UGA, a nonsense codon that causes termination of protein synthesis.

Upstream: Sequences proceeding in the opposite direction from expression (in the 5' direction) or located on the 5’ side of a particular point.

Wobble hypothesis: A hypothesis which accounts for the ability of a tRNA to recognize more than one codon by unusual pairing with the third base of a codon.

YAC: Abbreviation for a yeast artifical chromosome, which is a fragment of a chromosome from one species with regions of yeast chromosomes that permit the

chromosome to be perpetuated in yeast. YACs are used in the human genome project to clone large segments of genes from human or other species.

Zoo blot: Use of Southern blotting to test the ability of a DNA probe from one species to hybridize with the DNA of a variety of other species.